Members of the RKD transcription factor family induce an egg cell-like gene expression program

Koszegi, D ; Johnston, A J ; Rutten, T ; Czihal, A ; Altschmied, L ; Kumlehn, J ; Wüst, S E J ; Kirioukhova, O ; Gheyselinck, J ; Grossniklaus, U ; Bäumlein, H

Abstract: In contrast to animals, the life cycle of higher plants alternates between a gamete-producing (gametophyte) and a spore-producing (sporophyte) generation. The female gametophyte of angiosperms consists of four distinct cell types, including two gametes, the egg and the central cell, which give rise to embryo and endosperm, respectively. Based on a combined subtractive hybridization and virtual subtraction approach in wheat (Triticum aestivum L.), we have isolated a class of transcription factors not found in animal genomes, the RKD (RWP-RK domain-containing) factors, which share a highly conserved RWP-RK domain. Single-cell RT-PCR revealed that the genes TaRKD1 and TaRKD2 are preferentially expressed in the egg cell of wheat. The Arabidopsis genome contains five RKD genes, at least two of them, AtRKD1 and AtRKD2, are preferentially expressed in the egg cell of Arabidopsis. Ectopic expression of the AtRKD1 and AtRKD2 genes induces cell proliferation and the expression of an egg cell marker. Analyses of RKD-induced proliferating cells exhibit a shift of gene expression towards an egg cell-like transcriptome. Promoters of selected RKD-induced genes were shown to be predominantly active in the egg cell and can be activated by RKD in a transient protoplast expression assay. The data show that egg cell-specific RKD factors control a transcriptional program, which is characteristic for plant egg cells.

DOI: https://doi.org/10.1111/j.1365-313X.2011.04592.x

Running title: RKD controls an egg transcriptional program

Members of the RKD transcription factor family induce an egg cell-like gene expression program

Dávid Kőszegi[1], Amal J. Johnston[1], Twan Rutten[1], Andreas Czihal[1], Lothar Altschmied[1], Jochen Kumlehn[1], Samuel E. J. Wüst[2,a], Olga Kirioukhova[1,2], Jacqueline Gheyselinck[2,b], Ueli Grossniklaus[2] and Helmut Bäumlein[1]

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Gatersleben, Germany

[2]Institute of Plant Biology & Zürich-Basel Plant Science Center, University of Zürich, 8008 Zürich, Switzerland

Corresponding author:
Helmut Bäumlein
Institute of Plant Genetics and Crop Plant Research
Corrensstrasse 3, D-06466 Gatersleben
Germany
E-MAIL: baumlein@ipk-gatersleben.de
TEL: +49 39482 5238
FAX: +49 39482 5500

[a]Present address: Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

[b]Present address: Department of Plant Molecular Biology, University of Lausanne, Switzerland

SUMMARY

In contrast to animals, the life cycle of higher plants alternates between a gamete-producing (gametophyte) and a spore-producing (sporophyte) generation. The angiosperm female gametophyte consists of four distinct cell types, including two gametes, the egg and the central cell, which give rise to embryo and endosperm, respectively. Based on a combined subtractive hybridization and virtual subtraction approach in wheat (*Triticum aestivum* L.) we have isolated a class of transcription factors not found in animal genomes, the RKD factors, which share a highly conserved RWP-RK domain. Single cell RT-PCR revealed that the genes *TaRKD1* and *TaRKD2* are preferentially expressed in the egg cell of wheat. The *Arabidopsis* genome contains five *RKD* genes, at least two of them, *AtRKD1* and *AtRKD2,* are preferentially expressed in the egg cell of *Arabidopsis*. Ectopic expression of the *AtRKD1* and *AtRKD2* genes induces cell proliferation and the expression of an egg cell marker. Analyses of RKD-induced proliferating cells exhibit a shift of gene expression towards an egg cell-like transcriptome. Promoters of selected RKD-induced genes were shown to be predominantly active in the egg cell and can be activated by RKD in a transient protoplast expression assay. The data show that egg cell-specific RKD factors control a transcriptional program, which is characteristic for plant egg cells.

Keywords:

Wheat*, Arabidopsis*, female gametophyte, egg cell, RKD, transcription factor, transcriptome

INTRODUCTION

A typical plant life cycle comprises the alternation between a gamtophytic and sporophytic generation. The phylogeny of land plants is characterized by an evolutionary trend towards gametophyte reduction, as it has been described by Wilhelm Hofmeister more than a century ago (Hofmeister, 1851). In angiosperms the female gametophyte, the embryo sac, is strongly reduced and deeply embedded in sporophytic tissue. It originates from a diploid megaspore mother cell which undergoes meiosis. Of the resulting tetrad of haploid megaspores a single cell survives and develops into a seven-celled embryo sac. Within the embryo sac, the haploid egg cell and the diploid central cell are fertilized independently and give rise to a diploid embryo and triploid endosperm, respectively. This unique double fertilization event is a hallmark of angiosperm sexual reproduction (Grossniklaus and Schneitz, 1998; Yadegari and Drews, 2004).

Differentiation of the egg cell in the female gametophyte is tightly controlled, although the underlying molecular mechanisms are far from being understood. Recently, an auxin gradient was identified as an essential factor involved in the control of cell specification in the female gametophyte (Pagnussat *et al.*, 2009). Based on cytological observations and the analysis of mutant phenotypes in *Arabidopsis*, it has been proposed that the positioning of the nuclei within the female gametophyte is important for cell specification (Moore *et al.*, 1997; Pagnussat *et al.*, 2007; Webb and Gunning, 1990). In maize, small ubiquitin-related modifier-like proteins (diSUMO) have been shown to be involved in the segregation and positioning of nuclei during female gametophyte development (Srilunchang *et al.*, 2010). In the *indeterminate gametophyte1* (*ig1*) mutant of maize additional mitoses occur in the embryo sac, leading to the formation of supernumerary egg and central cells (Evans, 2007). In *Arabidopsis* several mutant collections affecting the development of the female gametophyte have been described (Christensen *et al.*, 1998; Pagnussat *et al.*, 2005). At least five *Arabidopsis* genes are known to control egg cell fate. Mutants in general splice factors like LACHESIS, CLOTHO/GFA1, and ATROPOS lead to the ectopic expression of an egg cell marker (Coury *et al.*, 2007; Gross-Hardt *et al.*, 2007; Johnston *et al.*, 2008; Moll *et al.*, 2008; Moore *et al.*, 1997). Moreover, the *eostre* mutant was found to cause ectopic expression of the homeodomain transcription factor BEL1, leading to a loss of synergid cell fate and the differentiation of an additional egg cell in *Arabidopsis*. Finally, the RETINOBLASTOMA RELATED (RBR) mutant, affecting the homolog of the animal

retinoblastoma tumor suppressor gene (*Rb*), functions as a negative regulator of gametophytic cell proliferation and differentiation (Ebel *et al.*, 2004; Johnston *et al.*, 2008; Johnston et al., 2010).

Molecular approaches such as (1) differential gene expression between wild-type and mutant ovules lacking a functional embryo sac (Johnston *et al.*, 2007; Jones-Rhoades *et al.*, 2007; Steffen *et al.*, 2007; Yu *et al.*, 2005); (2) large-scale sequencing of sequence tags, from egg cell cDNA libraries of maize (Cordts *et al.*, 2001; Le *et al.*, 2005; Yang *et al.*, 2006) and wheat (Kumlehn *et al.*, 2001; Sprunck *et al.*, 2005); and (3) microarray expression analysis of laser-dissected gametophytic cell types of *Arabidopsis* (Wuest *et al.*, 2010), have been used to identify additional components of a female gametophytic regulatory network. Despite these large-scale approaches, the molecular mechanisms of cell fate determination are largely unknown. However, they are of great interest not only from a developmental point of view but also for the engineering of apomixis, where sporophytic cells in the ovule initiate the formation of unreduced gametophytes or directly differentiate into embryos to produce clonal offspring (Koltunow and Grossniklaus, 2003).

Here, we report the functional characterization of a novel subclass of transcription factors of wheat and *Arabidopsis*. Based on their shared RWP-RK domain Schauser *et al.* (2005) named these factors RKD for RWP-RK domain-containing. The data suggest that members of the RKD family function as regulators of an egg cell related gene expression program.

RESULTS

Isolation of genes preferentially expressed in the wheat egg cell

To target molecular basis of egg cell-identity and development, a cDNA-library has been established from wheat egg cells and used in a combined hybridization and virtual subtraction approach to identify genes preferentially expressed in this cell type (Kumlehn et al., 2001). Firstly, clones carrying cDNAs of ubiquitously expressed genes were eliminated by hybridization with total cDNA derived from green leaves. 1139 non-hybridizing clones were sequenced and resulted in 1297 high-qualitiy sequences with an average sequence length of 354 bp. For further analysis, these clones were combined with 1094 EST sequences randomly chosen from the non-enriched clone pool (Kumlehn et al., 2001). Clustering of this dataset of 2391 ESTs

using the MIRA software (Chevreux et al., 2004) led to 849 unique sequences. Secondly, based on the notion that most cDNA libraries are made from tissues or plant organs in which egg cells and their transcripts are highly diluted or not present, the analysis was focused on 125 unique sequences which did not show any significant sequence similarity to more than one million publicly available wheat ESTs (Genebank release 171, May 2009, BLASTN score < 100). These sequences should represent transcripts which are either exceedingly rare and therefore potentially egg cell-specific or which represent non-plant contaminations introduced by the PCR-based construction of the egg cell cDNA library. The latter could be excluded for 22 of the 125 unique sequences, which encode plant genes as demonstrated by the fact that they show a significant sequence similarity (BLASTX score > 100) to rice proteins (MSU version 6.0, http://rice.plantbiology.mse.edu). After the two subtraction steps, three EST contigs (c10, c12, c413) were chosen to validate the approach.

To demonstrate the preferential egg cell expression of the three candidates, RT-PCR experiments were performed using RNA from anthers, carpels, leaves, stem and root, as well as from egg and central cells (Figure 1A). The three genes are neither detectably expressed in the above-mentioned tissues nor in the central cell. However, all three genes, designated egg cell factors (ECFs), are expressed in the egg cell, as detected by single cell RT-RCR on isolated cells of the embryo sac (Figure 1A).

The predicted amino acid sequence of the cDNA contig c10 exhibits sequence similarity to members of a class of plant transcription factors, which share a characteristic RWP-RK domain, preceded by a heptameric array of polar amino acids (Ferris and Goodengough, 1997; Schauser et al., 1999; Schauser et al., 2005; Lin and Goodenough, 2007). Based on protein size and domain sequence, the RWP-RK family can be divided into two subfamilies, the NIN-like proteins and the RKD proteins, which are clearly distinguishable in all available angiosperm genomes. Up to now, only members of the NIN-like subfamily were functionally characterized in *Lotus japonicus* (Schauser et al., 1999), *Pisum sativum* (Borisov et al., 2003) and *Medicago trunculata* (Marsh et al., 2007). The gene represented by the cDNA contig c10 represents a member of the RKD subfamily and was designated as *TaRKD*. The RKD gene family of wheat consists of at least four members as determined by Southern blot hybridization (Figure S1) and genomic sequencing (Figure S2). The exon-intron structure (Figure S2) was determined by comparison between the genomic and full-length cDNA sequences, obtained by 5`RACE. At the

transcript level the expression of two members, *TaRKD1* and *TaRKD2*, can be detected. No transcripts were found for the genes *TaRKD3* and *TaRKD4*.

The *AtRKD1* and *AtRKD2* genes of *Arabidopsis* are preferentially expressed in the egg cell

For a more detailed functional analysis of RKD gene we studied the homologous gene family in *Arabidopsis*. The *Arabidopsis* genome contains 14 *RWP-RK* genes (Schauser *et al.*, 2005), which can be subdivided into the NIN-like proteins and the RKD proteins (Figure S3). The *RKD* subfamily of *Arabidopsis* consists of at least five members: *AtRKD1* (At1g18790), *AtRKD2* (At1g74480), *AtRKD3* (At5g66990), *AtRKD4* (At5g53040) and *AtRKD5* (At4g35590).

Using quantitative real-time Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR), the highest transcript level of *AtRKD1* through *AtRKD4* was detected in ovules 2 days after emasculation (Figure 1C). In addition, faint *AtRKD1* and *AtRKD2* expression was found in flower buds in stages 1-11 (Smyth et al., 1990) and in siliques 2 days after pollination, but was not detectable in root, stem, leaf and anthers isolated from flowers in stages 11-13 (Figure 1C). Low amounts of *AtRKD3* transcripts were found in root, anthers and siliques (Figure 1C). Traces of *AtRKD4* transcripts were detected in leaves and bud and moderate transcript levels were found in anthers and siliques (Figure 1C). The relatively high amount of *AtRKD4* transcript in early embryos containing siliques is consistent with the observation that a mutation in this gene causes anatomical defects at the first zygote division (W. Lukowitz, pers. comm.). The more distantly related gene *AtRKD5* was found to be expressed in all tested tissues with the highest level in anthers (Figure 1C). The nodulin MtN3 family protein gene At5g40260, which was previously demonstrated to be preferentially expressed in both female and male gametophytes (Johnston *et al.*, 2007; Yu *et al.*, 2005) was used as control. The transcripts were found in buds, ovules, anthers and siliques (Figure 1C). Taken together, the genes *AtRKD1* to *AtRKD4* are mainly expressed in tissues containing the reproductive organs, while *AtRKD5* has a different profile, with expression in all examined samples.

The *AtRKD* expression profiles in female gametophytic tissues were analyzed by *in situ* hybridization experiments using gene-specific probes, excluding the conserved RWP-RK domain. Hybridization signals in the mature embryo sac were detected in the egg cell for *AtRKD2* and in the egg and synergid cells (the egg apparatus) for *AtRKD1* (Figure 2A and 2D). The specificity of the signal was checked using sense probes (Figure 2B and 2E). Both genes are

not detectably expressed at mitotic stages of the embryo sac. Consistent with the results described above for *TaRKD*, these data show that *AtRKD1* and *AtRKD2* are preferentially expressed in the egg cell or egg apparatus in the embryo sac.

Transgenic lines expressing the *uidA* gene (encodes β-glucuronidase, GUS), under the control of the *AtRKD1* and *AtRKD2* gene promoters, were generated. In at least ten independent *AtRKD2pro:GUS* transformants, GUS activity was only detected in the egg cell of the mature embryo sac, and similar results were obtained with *AtRKD1pro:GUS* (Figure 2C and 2F). Promoter activity could not be detected at earlier developmental stages of the female gametogenesis, but residual GUS activity was found in zygotes 24 hours after pollination, most likely due to persistence of the relatively stable GUS protein. No promoter activity was found in male gametophytes or in sporophytic tissues. These data demonstrate the preferential activity of the *AtRKD1* and *AtRKD2* promoters in the egg cell and support the qRT-PCR and *in situ* hybridization results, showing that these two genes are preferentially expressed in the egg cell.

To support this further we investigated whether *RKD* gene expression was de-regulated when egg cell differentiation was compromised. The RBR protein controls the differentiation and development of the female gametophyte (Ebel *et al.*, 2004; Johnston *et al.*, 2008), particularly the specification of its cell types (Johnston *et al.*, 2010), but also influences cell specification and differentiation in the sporophyte (Wildwater *et al.*, 2005; Wyrzykowska *et al.*, 2006). In the *rbr* mutant, mitotic divisions in the embryo sac are not arrested, and it undergoes excessive proliferation instead of differentiation, leading to a loss of egg cell specificity (Johnston et al., 2010). The *AtRKD1pro:GUS* transgene was specifically expressed in egg cells of wild-type but not *rbr* mutant embryo sacs, albeit in very few cases it appeared to be de-regulated (Figure 3). Thus, upon mis-specification of cell identity in the *rbr* mutant, *AtRKD1pro:GUS* activity is impaired, confirming its preferential egg cell expression.

RKD proteins are localized in the nucleus

Although it was suggested that RKD proteins function as nuclear transcription factors (Ferris and Goodenough, 1997; Schauser *et al.*, 1999), their subcellular localization remained unknown. Therefore, each of the *AtRKD1* through *AtRKD4* coding regions was fused in-frame to the coding region of the GREEN FLUORESCENCE PROTEIN (GFP) encoding gene (*d35Spro:AtRKD-GFP*), and transiently expressed in *Arabidopsis* protoplasts. As shown in

Figure S5 all four fusion proteins were localized in the nucleus of the cells, whereas the non-fused GFP gene product was also detectable in the cytoplasm. Thus, the nuclear localization of AtRKD proteins is consistent with their proposed role as transcriptional regulators.

Single and double mutants do not show an obvious phenotype

To gain insight into the function of the *AtRKD1* and *AtRKD2* genes expressed in the egg cell, we obtained T-DNA insertion lines from the Salk Institute Genomic Analysis Laboratory (Alonso *et al.*, 2003) and the GABI-KAT collection (Rosso *et al.*, 2003) (Figure S4). Homozygous mutant plants were identified for all the available alleles but they did not display obvious defects in either sporophytic or gametophytic tissues. Double mutants were generated combining the different *Atrkd1* and *Atrkd2* alleles (Table S1). None of the double mutants obvious morphological differences during female gametophyte development, most likely due to functional redundancy within the *AtRKD* gene family. Currently we aim for multiple mutants including alleles of *AtRKD3*, *AtRKD4* and *AtRKD5*.

Mis-expression of *AtRKD1* and *AtRKD2* leads to undifferentiated tissue formation

Gain of function experiments have been performed using ectopic expression of AtRKD1 through AtRKD4-cDNAs under the control of the double CaMV35S gene promoter. The ectopic expression of *AtRKD3* and *AtRKD4* produced no discernible phenotypes. In contrast, the expression of the *AtRKD1* and *AtRKD2* constructs causes severe distortions of plant growth including ectopic tissue proliferation (Figure S6). Similar growth distortions could be detected in plant lines with ectopic expression of *AtRKD1::GFP* and *AtRKD2::GFP* fusion genes demonstrating that the translational fusion to *GFP* did not interfere with the activity of these RKD proteins (Figure 4). The resulting tissue can be morphologically subdivided in organ differentiating green sections with large cells and in colourless sections with small proliferating cells. Remarkably, the RKD::GFP fusion proteins can only be detected in the nuclei of the proliferating small cells (Figure 4). No GFP signal is detectable in the differentiating green parts. Currently it is not clear what causes the loss of expression of the *RKD::GFP* constructs in the green part, however a gene silencing event might be a conceivable explanation. However, this provides the experimental advantage of an internal control to demonstrate the contrasting expression pattern and a clear correlation between the presence of both RKD::GFP fusion

proteins in the nucleus and the generation of the undifferentiated and proliferating tissue. Overall, these data indicate that mis-expression of *AtRKD1* and *AtRKD2* leads to the proliferation of cells that do not express differentiation markers such as chlorophyll.

Gametophytic markers are active in AtRKD-induced tissue

The distinct cytological features of the proliferating tissue prompted us to investigate the expression of gametophytic marker genes in the *AtRKD1::GFP* and *AtRKD2::GFP* expressing cells. We choose *At5g40260* and *At2g20070,* both known to be expressed in all cells of the embryo sac (Johnston *et al.*, 2007; Steffen *et al.*, 2007; Yu *et al.*, 2005), and the gene *At5g21030* which is preferentially expressed in the egg cell (Wuest *et al.*, 2010). RT-PCR shows gene expression in the white, proliferating tissue, whereas no expression was detected in green tissue (Figure 5). Seedlings at the cotyledon stage and auxin-induced, proliferating callus tissue are considered as sporophytic controls, whereas pistils prior to fertilization serve as gametophyte-containing control tissue. Expression was detectable in gametophyte-containing tissue only.

To specify this further, two gametophyte specific marker lines have been used. In the marker line ET1119 the egg cell is specifically labeled, whereas a construct consisting of the MEA gene promoter in front of the GUS reporter gene, controls a central cell specific expression (Gross-Hardt *et al.*, 2007; Figure S8). Here we demonstrate that the egg cel ET1119 marker becomes exclusively active in the small proliferating cells, whereas it is not expressed tissue consisting of the larger, chlorophyll containing cells (Figure 5). In contrast, the GUS reporter driven by the central cell specific MEA gene promoter is inactive both in the colourless and the green part (Figure 5). The data indicate that AtRKD factors confer sporophytic tissue the capability to adopt an egg cell -but not a central cell- related gene expression program.

Colorless tissue expresses a subset of egg cell transcriptome

For a more detailed analysis a genome-wide transcription profile of the RKD-induced proliferating tissue was determined using the Affymetrix® ATH1 array. Auxin-induced callus and two-week-old seedlings served as controls for proliferating cells and the sporophyte, respectively. Genes with less than three-fold increased signals were eliminated. The resulting 565 genes (Table S2, S3) were categorized according to biological functions (http://www.arabidopsis.org; (Berardini *et al.*, 2004). As expected for a highly specialized cell

type like egg cell, the majority of genes encodes proteins with unknown functions (Table 1). Hierarchical agglomerative sample clustering was applied to compare the global features of the transcriptomes of *AtRKD2-GFP*-induced and auxin-induced proliferating tissues. When the clustering was based on *AtRKD2*-induced genes (with a criterion of at least threefold up-regulation), *AtRKD2-GFP*-induced tissue was most similar to the egg cell, whereas auxin-induced callus tissue grouped with the root (Figure 6, Figure S9). These transcriptome data further support the suggestion that ectopic expression of *AtRKD2* induces the non-pigmented, proliferating cells to adopt transcriptome features of the egg cell.

Promoters of AtRKD2-GFP-induced genes are specifically active in the egg cell

Among the RKD2 induced genes described above, in total 107 genes (Table S3) with more than sevenfold induction and a p-value lower than 0.1 have been selected and further screened for low expression in various tissues using the GENEVESTIGATOR software (Zimmermann *et al.*, 2004). The activity of seven selected gene promoters (At1g53930, At1g56040, At1g60530, At1g66610, At3g12790, At3g62320, At4g04490) was tested using the chimaeric GFP::GUS reporter (Karimi *et al.*, 2002) in at least five independent transformants. In lines containing the constructs *At1g53930pro*:*GFP::GUS*, *At1g60530pro*:*GFP::GUS*, *At1g66610pro*:*GFP::GUS* and *At3g63320pro*:*GFP::GUS* the GFP signal was specifically detected in the egg cell (Figure 7). No signal was observed in the male gametophyte. Similar results were obtained using GUS as reporter (data not shown). The genes encode an ubiquitin-like protein (At1g53930*)*, a predicted nucleic acid binding protein (At3g63320), a dynamin-like protein (At1g60530) and a protein with similarity to Drosophila SEVEN IN ABSENTIA (At1g66610). The activity of the three other promoters was not detectable in either the male or female gametophyte. Thus, the identified gene promoters are components of an egg cell expression programme and represent new *Arabidopsis* egg cell markers in addition to those described before (Gross-Hardt *et al.*, 2007; Ingouff *et al.*, 2009; Steffen *et al.*, 2007). The data further support the above mentioned suggestion that AtRKD2 induces egg cell-expressed genes and initiates aspects of an egg cell regulatory program.

We further analyzed whether RKD factors are able to transiently activate egg cell expressed genes in an *Arabidopsis* protoplast system. Selected putative target promoters driving the GFP-GUS reporter were co-transformed with *AtRKD1* and *AtRKD2* both driven by the

double 35S promoter. The promoters of At1g60530 (dynamin), At3g63320 (nucleic acids binding protein) and At1g66610 (*SEVEN IN ABSENTIA*) were significantly up-regulated by fold changes of 5.67 (AtRKD1) and 4.18 (AtRKD2), 2.84 (AtRKD1) and 2.85 (AtRKD2), and 4.22 (AtRKD1) and 4.58 (AtRKD2), respectively (Figure 7). The At1g53930 (ubiquitin) gene promoter activity was neither induced by AtRKD1 (0.67) nor by AtRKD2 (0.64). The results demonstrate that AtRKD1 and AtRKD2 can transiently activate promoters of egg cell expressed genes. This is -in addition to the above described nuclear localization- in agreement with the previously suggested role of RKD as transcription factors (Ferris and Goodenough, 1997; Schauser et al., 1999)

DISCUSSION

The egg cell plays a key role in the life cycle of all higher organisms. Fertilization of the egg cell marks the transition between the gametophytic and the sporophytic generation in the life cycle of plants. Here we report the isolation and functional characterization of members of a transcription factor subfamily, designated as RKD factors. The wheat genes *TaRKD1* and *TaRKD2* are preferentially expressed in the egg cell of the mature embryo sac. The *Arabidopsis* genes *AtRKD1* and *AtRKD2* are highly expressed in the egg apparatus and the egg cell, respectively and the ectopic expression of *AtRKD2* induces a subset of an egg cell transcriptome. Selected RKD induced gene promoters exhibit egg cell specific activity. The data strongly suggest that RKD factors act as transcription factors involved in the regulation of an egg cell transcriptional network as basis for egg cell specification and differentiation.

Gametophyte development originates from the functional megaspore. Three mitotic divisions lead to a syncytium of eight nuclei followed by cellularization and differentiation. It has been proposed that these processes depend on nuclear location and migration within cytoplasmic domains (Brown and Lemmon, 1992). Regulatory proteins like IG1 and RBR of maize and *Arabidopsis*, respectively, are involved in the control of cell proliferation. Mutations in the corresponding genes lead to supernumerary nuclei, which are mis-positioned within the embryo sac, and eventually to the mis-specification of female gametophytic cells (Evans, 2007; Johnston *et al.*, 2008). Mis-specification of gametophytic cells was also observed in the *eostre* mutant of *Arabidopsis.* Here a BLH1-KNAT3 complex was shown to be involved in the switch from synergid to egg cell identity (Pagnussat *et al.*, 2007). An analogous interplay between cell

proliferation and differentiation has been proposed for the development of the male gametophyte. Here, *DUO1* is required for the division of sperm precursor cells as well as for promoting their differentiation into functional sperm cells (Brownfield *et al.*, 2009). Proposing analogous developmental processes in male and female gamete formation, it is well conceivable that RKD factors, in addition to or in cooperation with the abovementioned factors, may play a similar role in connecting cell proliferation and cellular differentiation programs during megagametogenesis. This hypothesis is supported by the finding that *AtRKD1* and *AtRKD2* have egg cell-specific functions in gene regulation.

The described RKD factors of wheat and *Arabidopsis* exhibit sequence similarity to other plant proteins containing the conserved RWP-RK domain, including 13 RWP-RK genes in the genome of *Chlamydomonas reinhardtii* (Riano-Pachon *et al.*, 2008). Remarkably, one of these gene products, MINUS DOMINANCE (MID*),* has been described to be necessary and sufficient for the development of minus gametes in this green algae. Consistent with the proposed function of MID in *Chlamydomonas*, we propose that RKD factors are involved in the control of egg cell functions, such as the differentiation between gametes and accessory, non-gametic (somatic) cells of the female gametophyte. Such a separation between germ line and soma in a gametophytic organism is best known in *Volvox carteri* (Tam and Kirk, 1991) but occurs in all gametophytes, even in so highly reduced ones as those of the angiosperms. A male-specific RKD-like gene has been isolated from the oogamous volvocacean species *Pleodorina starrii* (Nozaki *et al.*, 2006). This gene encodes a protein abundant in sperm nuclei and is only present in male genomes, suggesting a role in male gametogenesis. Sequence similarity, genomic occurrence and induction under nitrogen deprivation suggest that in the *Volvocaceae* family males have evolved from the dominant isogametic mating type (Nozaki *et al.*, 2006). Together, this proposes a high phylogenetic conservation of the gamete-related function of RKD factors.

Homology searches with the highly conserved RWP-RK motif reveal that animal genomes lack RKD homologues. This suggests a function of RKD proteins in a plant-specific process, as for instance a process required for the gametophytic generation. Both, in plant and animal reproductive processes, one cell is selected to undergo meiosis. This cell is called megaspore mother cell in plants and oocyte in animals. However, both kingdoms differ greatly in further processes of gamete differentiation. The surviving meiotic product of animals does not divide further and directly differentiates into the egg. As a plant specific hallmark of

reproduction, the surviving meiotic product, the functional megaspore, undergoes further mitotic divisions to generate the gamete producing gametophyte. Consistent with the proposed function of MID in *Chlamydomonas*, we suggest that plant RKD factors are involved in egg cell differentation form somatic gametophytic cells. This developmental step requires mechanisms to halt nuclear proliferation, to specify the gametes, and to distinguish them from the non-gametic, accessory cells of the embryo sac. Such a separation between gametic and non-gametic cells is required in all multicellular gametophytes, even including the highly reduced ones of the angiosperms. Thus, these conserved plant specific processes might require RKD functions, also explaining their absence from animal genomes.

In summary, predominant expression in egg cells of wheat and *Arabidopsis*, the induction of an egg cell-like transcriptome, egg cell activity and transient regulation of induced promoters, plant specific occurrence and phylogenetic conservation lead to the suggestion that RKD transcription factors of plants are involved in the regulation of female gamete development and capable to induce a subset of an egg cell transcription profile in sporophytic cells, causing a reprogramming process. The latter is not unlike examples in animals, where the expression of cell type-specific combinations of a few transcription factors can reprogram differentiated cells into a desired cell type, e.g. induced pluripotent stem cells (Yamanaka, 2008) or insulin-producing β-cells (Zhong *et al.*, 2008). Identification and analysis of downstream genes of the RKD factors should provide insights into the mechanisms controlling egg cell development. These studies will allow the identification and functional characterization of gene regulatory networks that operate during the specification and differentiation of this important cell type of the embryo sac and might provide tools to manipulate parthenogenetic processes as a component of apomictic reproduction.

EXPERIMENTAL PROCEDURES

Plant material

Arabidopsis thaliana (accession Columbia-0) was used; plants were grown on potting substrate 2 (Klasmann-Deilmann, Germany) at 23°C and 40% humidity with a light/dark cycle of 16 and 8 hours, respectively.

Cloning methods

Standard molecular techniques including Southern hybridization were performed as described (Sambrook *et al.*, 1989) and the GATEWAY™ technology (Invitrogen) was applied according to the manufacturer's protocol. *Escherichia coli* strain DH5α was used in routine cloning work. Oligonucleotides were obtained from Metabion AG (Martinsried, Germany) or Invitrogen (Karlsruhe, Germany).

Plant transformation

T-DNA constructs were first introduced into the *Agrobacterium tumefaciens* strain GV2260 by freeze-thaw transformation (Chen *et al.*, 1994). *Arabidopsis* was transformed using the floral-dip method (Clough and Bent, 1998).

5`RACE of wheat *TaRKD1* and *TaRKD2*

Non-fertilized egg cells of *aestivum*-Salmon wheat were isolated from emasculated spikes largely following a procedure described previously for fertilized wheat egg cells (Kumlehn et al. 1998). Further experimental details are given as supplements.

RT-PCR and qRT-PCR

RNA was isolated from different tissues using the Biomol solution (Biomol, Germany) according to the protocol provided by the supplier. 1 µg of RNA was used for cDNA synthesis after DNaseI treatment (2,5 units) (Roche, Germany), by RevertAid™ H Minus M-MuLV Reverse Transcriptase (MBI Fermentas, Germany) at 42°C for 60 minutes. Detailed PCR conditions and used primer are given as supplements.

***In situ* hybridization**

Inflorescences were embedded in paraplast following a published protocol (Kerk *et al.*, 2003). Gene-specific fragments were cloned (see primers below) into the pCRII-TOPO vector (Invitrogen, USA) following the protocol of the manufacturer. These plasmids were used as templates for generating digoxygenin-UTP-labeled riboprobes by run-off transcription using T7 and SP6 RNA polymerases according to the manufacturer´s protocol (Roche Diagnostics, Switzerland). *In situ* hybridization was performed on 8-10 µm semi-thin paraffin sections as

previously described (Vielle-Calzada *et al.*, 1999). The following primers were used to construct the *in situ* probes: AtRKD1-forward, AtRKD1-reverse, AtRKD2-forward, AtRKD2-reverse.

Promoter:GUS reporter fusion constructs

A 1,303 bp fragment upstream of the *AtRKD1* start codon was cloned into the pMDC163 vector (Curtis and Grossniklaus, 2003). Fragments of 522 bp, 1,315 bp and 436 bp length upstream of the start codon of the genes *AtRKD2*, *AtRKD3* and *AtRKD4*, respectively, were cloned into the pBIN19 vector (Bevan, 1984), carrying an intron containing *uidA* gene encoding GUS. The used primer oligonucleotides are given as supplements. For GUS detection the harvested plant material was vacuum-infiltrated, incubated overnight at 37°C in the *GUS* staining solution solution (Biosynth, Switzerland) and cleared for 15 min in 20% lactic acid and 20% glycerol and analyzed with a light microscope (Axioplan, Zeiss, Germany). Promoters of putative target genes were cloned using GATEWAY technology into the plasmid pKGWFS7.0 containing the chimaeric GFP::GUS reporter (Karimi et al, 2002). GFP signals were localized *in vivo* using a confocal laser-scanning microscope (Zeiss, Germany). The GFP fluorophore was excited at a wavelength of 488 nm by an argon laser and detected at wavelengths between 505 nm and 520 nm.

Transient expression in protoplasts for subcellular localization

RKD coding regions from the start codon through the last amino acid codon were PCR amplified and integrated into the GATWAY destination vector pMDC84 (Curtis and Grossniklaus, 2003). These constructs were used for transient expression in tissue culture-derived *Arabidopsis* protoplasts as described previously (Ivanov et al., 2008). GFP signals were localized *in vivo* using a confocal laser-scanning microscope (Zeiss, Germany).

Transient expression of promoter:reporter constructs in protoplasts

The promoters of genes At1g53930, At1g60530, At1g66610 and At3g63320 were cloned into pKGWFS7.0 plasmid (Karimi et al, 2002). AtRKD1 and AtRKD2 amplicons were integrated into the GATEWAY destination vector pMDC32 (Curtis and Grossniklaus, 2003). For protoplast transformation, aliquots of 330 µl were heat-shocked (42°C for 5 minutes) before plasmid DNA (5 µg of each plasmid) and carrier DNA (160 µg of calf thymus DNA) were

added. PEG 6000 (final concentration 20%) was used to induce DNA uptake. After 72 hours incubation in the dark at room temperature, protoplasts were harvested, and the GUS activity was determined by a fluorimetric assay (Jefferson, 1987) using the GUS-LightTM Kit (Tropix, Bedford, USA). An AtUBQ10pro:LUC plasmid was used as a normalization control for transformation efficiency. Each experiment was repeated three times and the average values were calculated.

Characterization of the *Arabidopsis* T-DNA lines

To identify plants with the T-DNA insertion in *AtRKD1* and *AtRKD2,* PCR analyses were performed. Allele-specific PCR reactions were performed to confirm the T-DNA insertion sites using primers for GABI lines (o8409) and for SALK lines (Rba3 or LBb1). Gene-specific primers are given as supplement.

Array hybridization

Total RNA was extracted from AtRKD1-GFP-, AtRKD2-GFP-, and auxin-induced callus tissue and 14-days-old seedlings using Trizol reagent. The labelling and hybridization were performed by ATLAS Biolabs GmbH (Germany).

Signal calculation and sample clustering

To determine gene expression signals, Li-Wong expression indexes were calculated in the DNA-Chip Analyzer Software (dChip 2008, Li and Wong, 2001) using invariantset-normalization and the PM-only model. Follow-up analyses were performed in the statistical software "R" (Version 2.8.0, http://www.r-project.org/) and Bioconductor software packages (www.bioconductor.org). Hierarchical agglomerative sample clustering was performed using the pvclust-package for assessing the uncertainty of the clustering based on resampling (Suzuki and Shimodaira, 2006). Gametophyte-enriched genes were determined by comparing the cell-type-specific expression profiles with a large compendium of publicly available tissue/cell-type-specific expression profiles (Wuest *et al.*, 2010) including data from the Goldberg-Harada embryo compartment datasets (GSE12404 record in GEO, http://www.ncbi.nlm.nih.gov/gds) as used in Le *et al.*, (2010). Probe-set linear models on log2-dChip expression signals were fitted using the package "limma" (Smyth, 2004), and pair-wise contrasts of all other tissue/cell types against the cell-type

of interest were examined using an empirical Bayesian approach as implemented in the package. P-value adjustments were performed using the Bonferroni-Holm method, and a maximum p-value of 0.01 between all contrasts examined was considered significant.

ACKNOWLEDGEMENTS

We would like to acknowledge the excellent technical assistance of Elke Liemann, Sabine Skiebe and Alexandra Rech. We thank Anne Tewes for help with the transient expression assays, Annchristin Zierold for help with RNA amplification from isolated egg cells and Maria Mildner, Eszter Kapusi, Isolde Saalbach and Urs Hähnel for discussions and support. This work was primarily supported by the ApoTool Project (FP5) to H. B. and the German Ministry for Education and Research in the frame of the GABI-SEEDII project and Deutsche Forschungsgemeinschaft (DFG) grant (No.: BA1235/11-1) to D.K. and H.B. and by the core funds of IPK-Gatersleben to H.B.. A.J.J. is a Humboldt Fellow and acknowledges a short-term fellowship based at H.B.'s lab from the European Molecular Biology Organization (EMBO). U.G acknowledges grants from the ApoTool Project (FP5), University of Zürich and the Swiss National Science Foundation (3100-064061 and 3100-112489).

Table Legend

Table 1. Classification of AtRKD2-induced genes with an at least 3-fold change in expression as compared to auxin callus. The categorization is based on the current annotation of The *Arabidopsis* Information Resource at http://www.arabidopsis.org.

Table 1.

Functional Category	**Gene count (%)**
other cellular processes	21.43
other metabolic processes	20.37
unknown biological processes	16.99
protein metabolism	6.66
response to abiotic or biotic stimulus	6.18
Transcription	5.60
response to stress	5.41
developmental processes	4.92
other biological processes	4.05
cell organization and biogenesis	3.09
Transport	2.32
DNA or RNA metabolism	2.22
electron transport or energy pathways	0.58
signal transduction	0.19
∑	100.00

Figure Legends

Figure 1. Genes with preferential expression in plant egg cells.

(A) Wheat single cell RT-PCR analysis of *TaRKD*, *ECF2* and *ECF3* genes in different gametophytic and sporophytic cell types. The constitutively expressed gene for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as control.

(B) Alignment of RWP-RK domains of wheat TaRKD, *Arabidopsis* AtRKD1-5 and *Chlamydomonas* MID protein. Identical and similar amino acids are given with a black and grey background, respectively. Stars indicate the heptad repeats of large hydrophobic amino acid side chains. The arrowhead indicates the position of the K124 residue of MID, known to be essential for the normal function of the protein (Ferris and Goodenough, 1997). The alignment was obtained using AlignX software (Invitrogen, CA, USA).
(C) qRT-PCR analysis of *Arabidopsis AtRKD1, AtRKD2, AtRKD3, AtRKD4, AtRKD5* and the nodulin MtN3 family protein *(At5g40260)* genes in roots (R), stems (St), leaves (L), flower buds at stages 1 to 11 (B), ovules 2 days after emasculation (O), anthers from flowers at stages 11-13 (A) and siliques 1-2 days after pollination (Si). Each experiment was repeated three times and in each repetition three independent PCR reactions were carried out.

Figure 2. *AtRKD1* and *AtRKD2* are expressed in the egg apparatus and the egg cell, respectively. **(A)** to **(F)** *AtRKD* transcript localization and detection of promoter activity. *In situ* hybridization with gene-specific probes for **(A)** *AtRKD1* antisense, **(B)** *AtRKD1* sense control, **(D)** *AtRKD2* antisense, and **(E)** *AtRKD2* sense control. Localization of promoter:GUS activity in plants transformed with **(C)** *AtRKD1pro:GUS* and **(F)** *AtRKD2pro:GUS* constructs. Egg cells and synergids are labelled with red and green arrowheads, respectively. Black arrowheads indicate the GUS signal in the egg cell. Bars, 20 μm.

Figure 3. Promoter activity in embryo sacs of the wild type and *rbr* mutant. (*A*) The *AtRKDpro:GUS* construct is active in the egg cell of the wild type. (*B*) The majority of the *rbr* mutant embryo sacs, in which egg cells are not specified (Johnston *et al.,* 2008) did not express egg cell-specific *AtRKD1pro:GUS*. Black arrows mark the proliferating cells. (*C-E*) In some cases the construct is mis-expressed in *rbr* mutant embryo sacs, consistent with the egg cell being mis-specified in rare *rbr* mutant gametophytes: mis-expression in two egg cell-like structures (*C*); mis-expression in an egg apparatus-like structure (*D*); mis-expression throughout the embryo sac (*E*). Histogram of phenotypic classes in the *rbr-3* allele. Note that *AtRKD1pro:GUS* was heterozygous. Total counts for *RBR*/*RBR* and *RBR*/*rbr* were 196 and 228, respectively. Class I: GUS staining in egg cell; class II: absence of GUS staining in egg cell; class III: mis-expression of GUS either in the egg cell or in several embryo sac cells. Black

columns are for the wild type, grey bars for the *rbr* mutant. Bars, 30 µm.

Figure 4. Phenotype of *AtRKD-GFP* over-expressing tissue and localization of RKD-GFP fusion proteins. **(A)** to **(E)** Ectopic expression of *AtRKD1-GFP* and **(F)** to **(J)** *AtRKD2-GFP* leads to the generation of dimorphic tissue with a colorless, proliferating part and a differentiating, green part. **(A-B)** and **(F-G)** Brightfield images, **(C)** and **(H)** detection of chloroplasts, **(D)** and **(I)** visualization of AtRKD1- and AtRKD2-GFP fusion proteins. **(E)** and **(J)** Overlay of the images B-D and G-I. The fusion proteins are exclusively detected in the nuclei of small, proliferating cells [**(B)** and **(I)**] but absent in large, differentiating cells with chloroplasts, which exhibit red fluorescence [**(C)** and **(F)**]. Bars, 1 mm in **(A)** and **(F),** 20 µm in **(B)**, **(C), (D), (E)** and 50 µm in **(G), (H), (I), (J)**.

Figure 5. Activity of egg cell and central cell markers in *AtRKD-GFP* over-expressing tissue. The egg cell-specific maker ET1119 is exclusively expressed in the proliferating, colorless part induced by the AtRKD2-GFP fusion protein **(A)**, but not in the differentiating, green part **(B)**, although remnants of expressing tissue can be seen. **(C)** and **(D)** In contrast, the central cell-specific *MEApro:GUS* marker is inactive in both parts of the tissue. Bars, 1 mm in **(C)**, **(D)** and 2 mm in **(A)**, **(B)**. Expression of gametophytic marker genes in AtRKD1- and AtRKD2-induced tissue. **(E)** RT-PCR for the gene *At5g40260* (encoding a nodulin-like protein) (Johnston et al., 2007), **(F)** quantitative RT-PCR for the gene *At2g20070* (DD33) (Steffen et al., 2007) and **(G)** for *At5g21030* (Wuest et al., 2010).

Figure 6. Comparative transcriptome analysis between cells expressing AtRKD2-GFP and sporophytic and gametophytic tissues and cell types. Hierarchical agglomerative sample clustering based on euclidean distances was applied for genes upregulated in AtRKD2-GFP-induced proliferating tissue. The sample clustering is based on genes that are at least three-fold upregulated in callus tissue when compared to control callus (a total of 490 genes). Note that AtRKD2-GFP callus and egg cell cluster together (red arrow), whereas auxin-induced control callus group with root. Node labels denote bootstrap support from 10,000 replications, with red numbers denoting bootstrap probabilities and green numbers denoting approximately unbiased

probability values.

Figure 7. Promoter activity of AtRKD2-induced genes in egg cells and promoter activities with and without AtRKD1 or AtRKD2 in a transient *Arabidopsis* protoplast system.
(A) to **(D)** Promoter activity using GFP reporter protein, detected by fluorescence microscopy/laser scanning microscopy (Zeiss, Germany). The GFP fluorophore was excited at 488 nm by an argon laser and detected between 505 nm and 520 nm. Only merged images are shown for **(A)** *At1g53930pro:GFP*, **(B)** *At1g60530pro:GFP*, **(C)** *At3g62320pro:GFP*, and **(D)** *At1g66610pro:GFP*. Bars 20 µm. **(E)** The promoter activities are given in fold change in the presence of either AtRKD1 or AtRKD2 compared to the control. Stars indicate significant differences calculated by the Student t-test. The GUS activity was measured 3 days after transformation. Each experiment was repeated three times.

LITERATURE

Alexander, M.P. (1969) Differential staining of aborted and nonaborted pollen. *Stain. Technol.*, **44**, 117-122.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C. and Ecker, J.R. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653-657.

Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L.A., Yoon, J., Doyle, A., Lander, G., Moseyko, N., Yoo, D., Xu, I., Zoeckler, B., Montoya, M., Miller, N., Weems, D. and Rhee, S.Y. (2004) Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiol.*, **135**, 745-755.

Bevan, M. (1984) Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Res.*, **12**, 8711-8721.

Borisov, A.Y., Madsen, L.H., Tsyganov, V.E., Umehara, Y., Voroshilova, V.A., Batagov, A.O., Sandal, N., Mortensen, A., Schauser, L., Ellis, N., Tikhonovich, I.A. and Stougaard, J. (2003) The Sym35 gene required for root nodule development in pea is an ortholog of Nin from Lotus japonicus. *Plant Physiol.*, **131**, 1009-1017.

Brown, R.C. and Lemmon, B.C. (1992) Cytoplasmic domain: A model for spatial control of cytokinesis in reproductive cells of plants. *EMSA Bull.*, **22**, 48-53.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T. and Twell, D. (2009) A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genet.*, **5**, e1000430.

Biotechniques, **16**, 664-668, 670.

Chevreux, B., Pfister, T., Drescher, B., Driesel, A.J., Muller, W.E., Wetter, T. and Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.,* **14**, 1147-1159.

Christensen, C.A., Subramanian, S. and Drews, G.N. (1998) Identification of gametophytic mutations affecting female gametophyte development in Arabidopsis. *Dev. Biol.*, **202**, 136-151.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.*, **16**, 735-743.

Cordts, S., Bantin, J., Wittich, P.E., Kranz, E., Lorz, H. and Dresselhaus, T. (2001) *ZmES* genes encode peptides with structural homology to defensins and are specifically expressed in the female gametophyte of maize. *Plant J.*, **25**, 103-114.

Coury, D.A., Zhang, C., Ko, A., Skaggs, M.I., Christensen, C.A., Drews, G.N., Feldmann, K.A. and Yadegari, R. (2007) Segregation distortion in Arabidopsis *gametophytic factor 1* (*gfa1*) mutants is caused by a deficiency of an essential splicing factor. *Sex. Plant Reprod.*, **20**, 87–97.

Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.*, **133**, 462-469.

Ebel, C., Mariconti, L. and Gruissem, W. (2004) Plant retinoblastoma homologues control nuclear proliferation in the female gametophyte. *Nature*, **429**, 776-780.

Evans, M.M. (2007) The *indeterminate gametophyte1* gene of maize encodes a LOB domain protein required for embryo sac and leaf development. *Plant Cell*, **19**, 46-62.

Ferris, P.J. and Goodenough, U.W. (1997) Mating type in Chlamydomonas is specified by *mid*, the minus-dominance gene. *Genetics*, **146**, 859-869.

Gross-Hardt, R., Kagi, C., Baumann, N., Moore, J.M., Baskar, R., Gagliano, W.B., Jurgens, G. and Grossniklaus, U. (2007) *LACHESIS* restricts gametic cell fate in the female gametophyte of Arabidopsis. *PLoS Biol*, **5**, e47.

Grossniklaus, U. and Schneitz, K. (1998) The molecular and genetic basis of ovule and megagametophyte development. *Semin. Cell. Dev. Biol.*, **9**, 227-238.

Hofmeister, W. (1851) Entfaltung und Fruchtbildung höherer Kryptogamen (Moose, Farrn, Equisetaceen, Rhizocarpeen und Lycopodiaceen) und die Samenbildung der Coniferen. *Vergleichende Untersuchungen der Keimung.*

Ingouff, M., Sakata, T., Li, J., Sprunck, S., Dresselhaus, T. and Berger, F. (2009) The two male gametes share equal ability to fertilize the egg cell in *Arabidopsis thaliana*. *Curr. Biol.*, **19**, R19-20.

Ivanov, R., Tiedemann, J., Czihal, A., Schallau, A., Le, H. D., Mock, H.P., Claus, B., Tewes, A. and Bäumlein, H. (2008) EFFECTOR OF TRANSCRIPTION2 is involved in xylem differentiation and includes a functional DNA single strand cutting domain. *Dev. Biol*., **313**, 93-106.

Johnston, A.J., Kirioukhova, O., Barrell, P.J., Moore, J.M., Baskar, R., Grossniklaus, U. and Gruissem, W. (2010) Dosage-sensitive function of RETINOBLASTOMA-

RELATED and convergent epigenetic control are required during the Arabidopsis life cycle. *PLoS Genet.*, **6**, e1000988.

Johnston, A.J., Matveeva, E., Kirioukhova, O., Grossniklaus, U. and Gruissem, W. (2008) A dynamic reciprocal RBR-PRC2 regulatory circuit controls Arabidopsis gametophyte development. *Curr. Biol.*, **18**, 1680-1686.

Johnston, A.J., Meier, P., Gheyselinck, J., Wuest, S.E., Federer, M., Schlagenhauf, E., Becker, J.D. and Grossniklaus, U. (2007) Genetic subtraction profiling identifies genes essential for Arabidopsis reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. *Genome Biol.*, **8**, R204.

Jones-Rhoades, M.W., Borevitz, J.O. and Preuss, D. (2007) Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. *PLoS Genet.*, **3**, 1848-1861.

Karimi, M., Inze, D. and Depicker, A. (2002) GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci.*, **7**, 193-195.

Kerk, N.M., Ceserani, T., Tausta, S.L., Sussex, I.M. and Nelson, T.M. (2003) Laser capture microdissection of cells from plant tissues. *Plant Physiol.*, **132**, 27-35.

Koltunow, A.M. and Grossniklaus, U. (2003) Apomixis: a developmental perspective. *Annu. Rev. Plant Biol.*, **54**, 547-574.

Kumlehn, J., Kirik, V., Czihal, A., Altschmied, L., Matzk, F., Lörz, H. and Bäumlein, H. (2001) Parthenogenetic egg cells in wheat: Cellular and molecular studies. *Sex. Plant Reprod.*, **14**, 239-243.

Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G.N., Fischer, R.L., Okamuro, J.K., Harada, J.J. and Goldberg, R.B. (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *P Natl Acad Sci USA*, **107**, 8063-8070.

Le, Q., Gutierrez-Marcos, J.F., Costa, L.M., Meyer, S., Dickinson, H.G., Lorz, H., Kranz, E. and Scholten, S. (2005) Construction and screening of subtracted cDNA libraries from limited populations of plant cells: a comparative analysis of gene expression between maize egg cells and central cells. *Plant J.*, **44**, 167-178.

Li, C. and Wong, W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. U S A*, **98**, 31-36.

Li, C. and Wong, W.H. (2001) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol.*, **2**, R0032.

Lin, H. and Goodenough, U.W. (2007) Gametogenesis in the *Chlamydomonas reinhardtii* minus mating type is controlled by two genes, *MID* and *MTD1*. *Genetics*, **176**, 913-925.

Maheshwari, P. and Johri, B.M. (1950) Development of the embryo sac, embryo and endosperm in helixanthera ligustrina (wall.) dans. *Nature*, **165**, 978-979.

Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.R., Schultze, M., Ratet, P. and Oldroyd, G.E. (2007) *Medicago truncatula* NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol.*, **144**, 324-335.

Moll, C., von Lyncker, L., Zimmermann, S., Kagi, C., Baumann, N., Twell, D., Grossniklaus, U. and Gross-Hardt, R. (2008) CLO/GFA1 and ATO are novel regulators of gametic cell fate in plants. *Plant J.*, **56**, 913-921.

Moore, J.M., Calzada, J.P., Gagliano, W. and Grossniklaus, U. (1997) Genetic

characterization of hadad, a mutant disrupting female gametogenesis in *Arabidopsis thaliana*. *Cold Spring Harb. Symp. Quant. Biol.*, **62**, 35-47.

Nozaki, H., Mori, T., Misumi, O., Matsunaga, S. and Kuroiwa, T. (2006) Males evolved from the dominant isogametic mating type. *Curr. Biol.*, **16**, R1018-1020.

Pagnussat, G.C., Alandete-Saez, M., Bowman, J.L. and Sundaresan, V. (2009) Auxin-dependent patterning and gamete specification in the Arabidopsis female gametophyte. *Science*, **324**, 1684-1689.

Pagnussat, G.C., Yu, H.J., Ngo, Q.A., Rajani, S., Mayalagu, S., Johnson, C.S., Capron, A., Xie, L.F., Ye, D. and Sundaresan, V. (2005) Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development*, **132**, 603-614.

Pagnussat, G.C., Yu, H.J. and Sundaresan, V. (2007) Cell-fate switch of synergid to egg cell in Arabidopsis *eostre* mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene *BLH1*. *Plant Cell*, **19**, 3578-3592.

Riano-Pachon, D.M., Correa, L.G., Trejos-Espinosa, R. and Mueller-Roeber, B. (2008) Green transcription factors: a chlamydomonas overview. *Genetics*, **179**, 31-39.

Richert, J., Kranz, E., Lörz, H. and Dresselhaus, T. (1996) A reverse transcriptase polymerase chain reaction assay for gene expression studies at the single cell level. *Plant Sci*. **114**, 93-99.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K. and Weisshaar, B. (2003) An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol.*, **53**, 247-259.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, 2nd edn. In *Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.*

Schauser, L., Roussis, A., Stiller, J. and Stougaard, J. (1999) A plant regulator controlling development of symbiotic root nodules. *Nature*, **402**, 191-195.

Schauser, L., Wieloch, W. and Stougaard, J. (2005) Evolution of NIN-like proteins in Arabidopsis, rice, and *Lotus japonicus*. *J. Mol. Evol.*, **60**, 229-237.

Smyth, D.R., Bowman, J.L. and Meyerowitz, E.M. (1990) Early flower development in Arabidopsis. *Plant Cell*, **2**, 755-767.

Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, **3**, Article3.

Sprunck, S., Baumann, U., Edwards, K., Langridge, P. and Dresselhaus, T. (2005) The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). *Plant J.*, **41**, 660-672.

Srilunchang, K.O., Krohn, N.G. and Dresselhaus, T. (2010) DiSUMO-like DSUL is required for nuclei positioning, cell specification and viability during female gametophyte maturation in maize. *Development*, **137**, 333-345.

Steffen, J.G., Kang, I.H., Macfarlane, J. and Drews, G.N. (2007) Identification of genes expressed in the Arabidopsis female gametophyte. *Plant J.*, **51**, 281-292.

Suzuki, R. and Shimodaira, H. (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, **22**, 1540-1542.

Tam, L.W. and Kirk, D.L. (1991) The program for cellular differentiation in Volvox carteri as revealed by molecular analysis of development in a gonidialess/somatic regenerator mutant. *Development*, **112**, 571-580.

Vielle-Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A. and

Grossniklaus, U. (1999) Maintenance of genomic imprinting at the Arabidopsis *medea* locus requires zygotic DDM1 activity. *Genes Dev.*, **13**, 2971-2982.

Webb, M.C. and Gunning, B.E.S. (1990) Embryo sac development in *Arabidopsis thaliana*. 1. Megasporogenesis, including the microtubular cytoskeleton. *Sex. Plant Reprod.*, **3**, 244-256.

Wildwater, M., Campilho, A., Perez-Perez, J.M., Heidstra, R., Blilou, I., Korthout, H., Chatterjee, J., Mariconti, L., Gruissem, W. and Scheres, B. (2005) The *RETINOBLASTOMA-RELATED* gene regulates stem cell maintenance in Arabidopsis roots. *Cell*, **123**, 1337-1349.

Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenfuhrer, J., von Mering, C. and Grossniklaus, U. (2010) Arabidopsis Female Gametophyte Gene Expression Map Reveals Similarities between Plant and Animal Gametes. *Curr. Biol.*, **20**, 506-512.

Wyrzykowska, J., Schorderet, M., Pien, S., Gruissem, W. and Fleming, A.J. (2006) Induction of differentiation in the shoot apical meristem by transient overexpression of a retinoblastoma-related protein. *Plant Physiol.*, **141**, 1338-1348.

Yadegari, R. and Drews, G.N. (2004) Female gametophyte development. *Plant Cell*, **16 Suppl**, S133-141.

Yamanaka, S. (2008) Induction of pluripotent stem cells from mouse fibroblasts by four transcription factors. *Cell Prolif.*, **41 Suppl 1**, 51-56.

Yang, H., Kaur, N., Kiriakopolos, S. and McCormick, S. (2006) EST generation and analyses towards identifying female gametophyte-specific genes in *Zea mays* L. *Planta*, **224**, 1004-1014.

Yu, H.J., Hogan, P. and Sundaresan, V. (2005) Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling. *Plant Physiol.*, **139**, 1853-1869.

Zhong, L.P., Yang, X., Zhang, L., Wei, K.J., Pan, H.Y., Zhou, X.J., Li, J., Chen, W.T. and Zhang, Z.Y. (2008) Overexpression of insulin-like growth factor binding protein 3 in oral squamous cell carcinoma. *Oncol. Rep.*, **20**, 1441-1447.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol.*, **136**, 2621-2632.

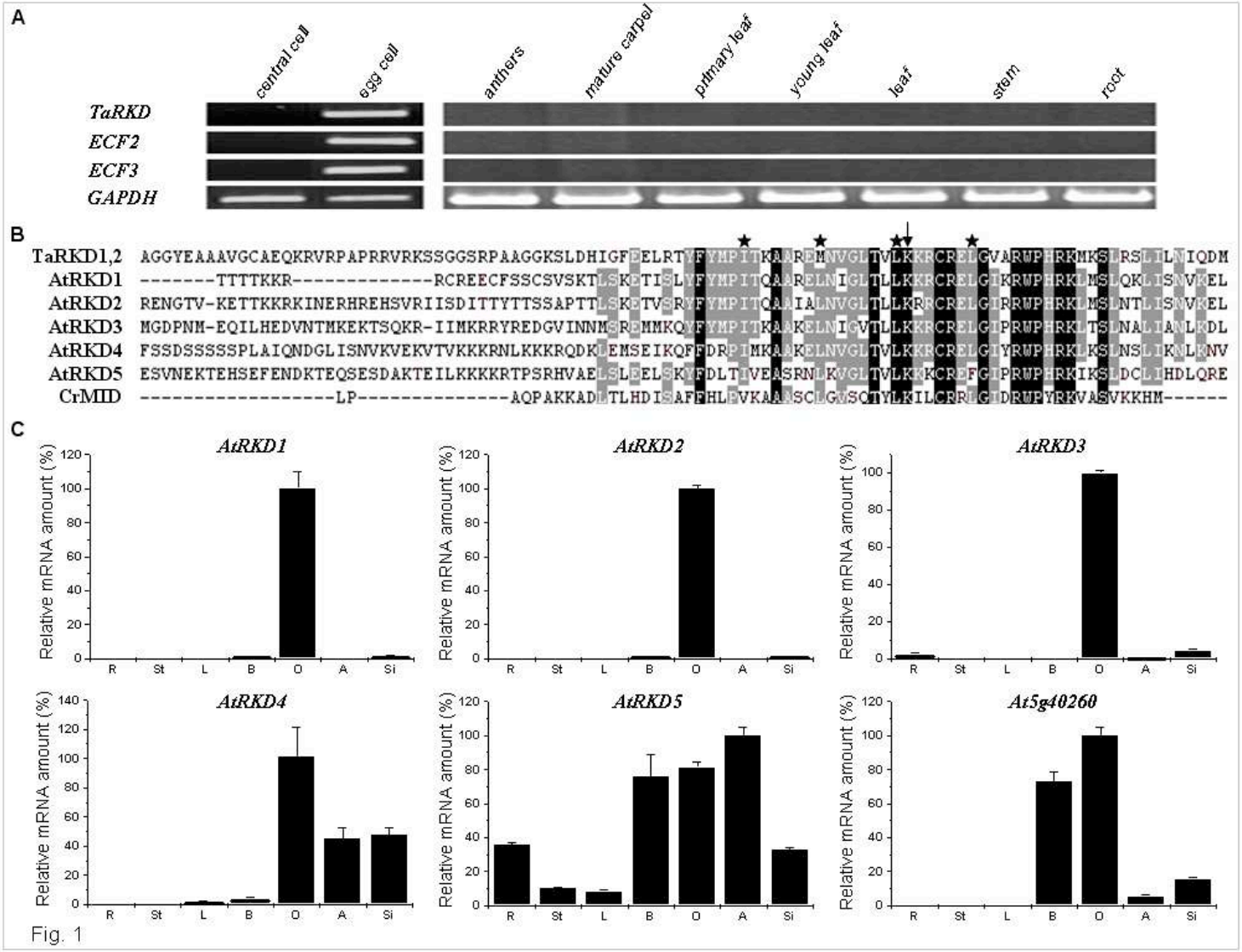

A
central cell
egg cell
anthers
mature carpel
primary leaf
young leaf
leaf
stem
root
TaRKD
ECF2
ECF3
GAPDH
B
TaRKD1,2
AtRKD1
AtRKD2
AtRKD3
AtRKD4
AtRKD5
CrMID
C
AtRKD1
AtRKD2
AtRKD3
AtRKD4
AtRKD5
At5g40260
Relative mRNA amount (%)
R
St
L
B
O
A
Si

Fig. 1

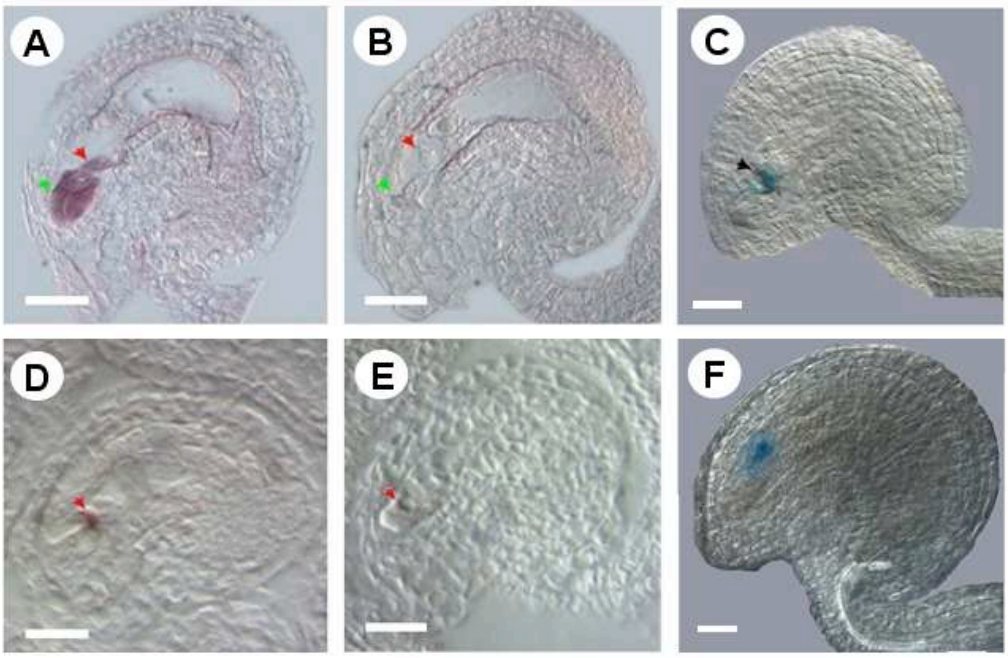


Fig. 2

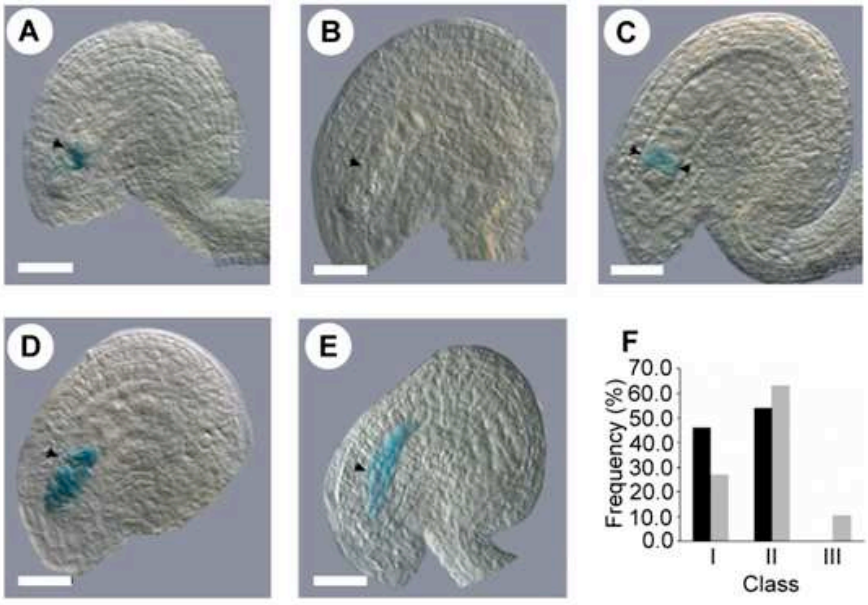


Fig. 3

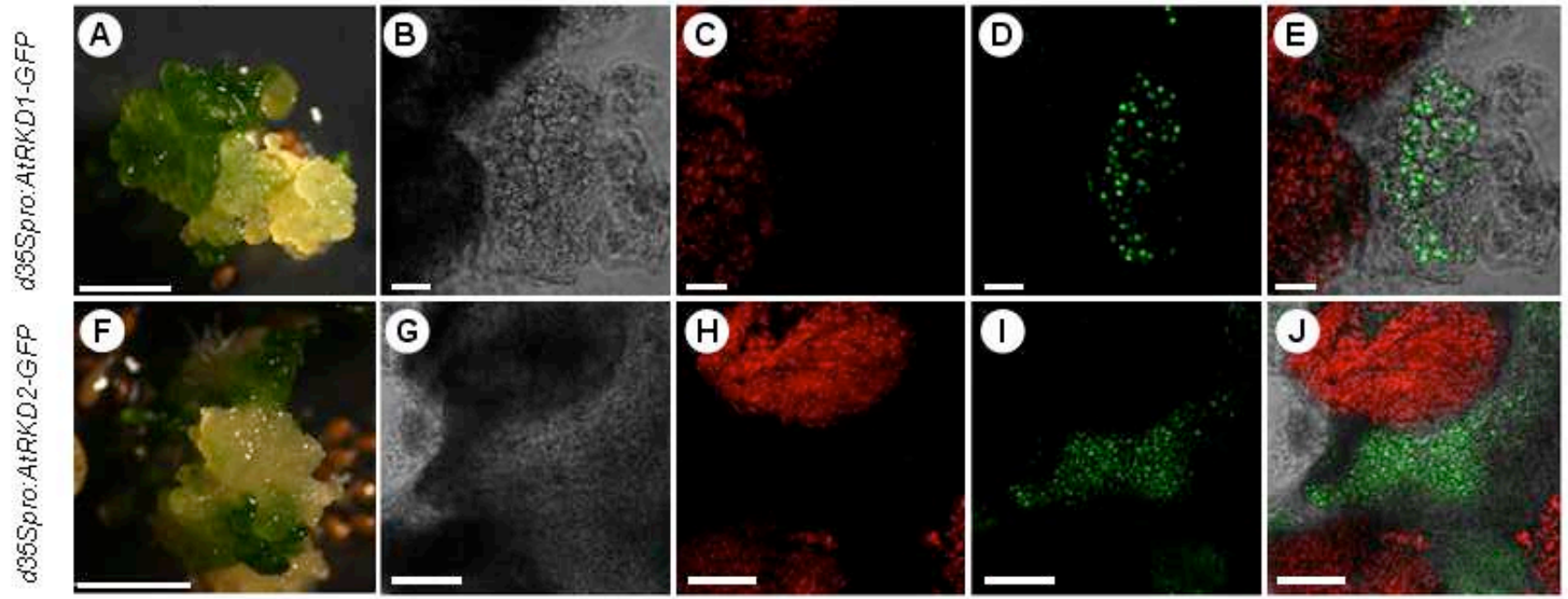


Fig. 4

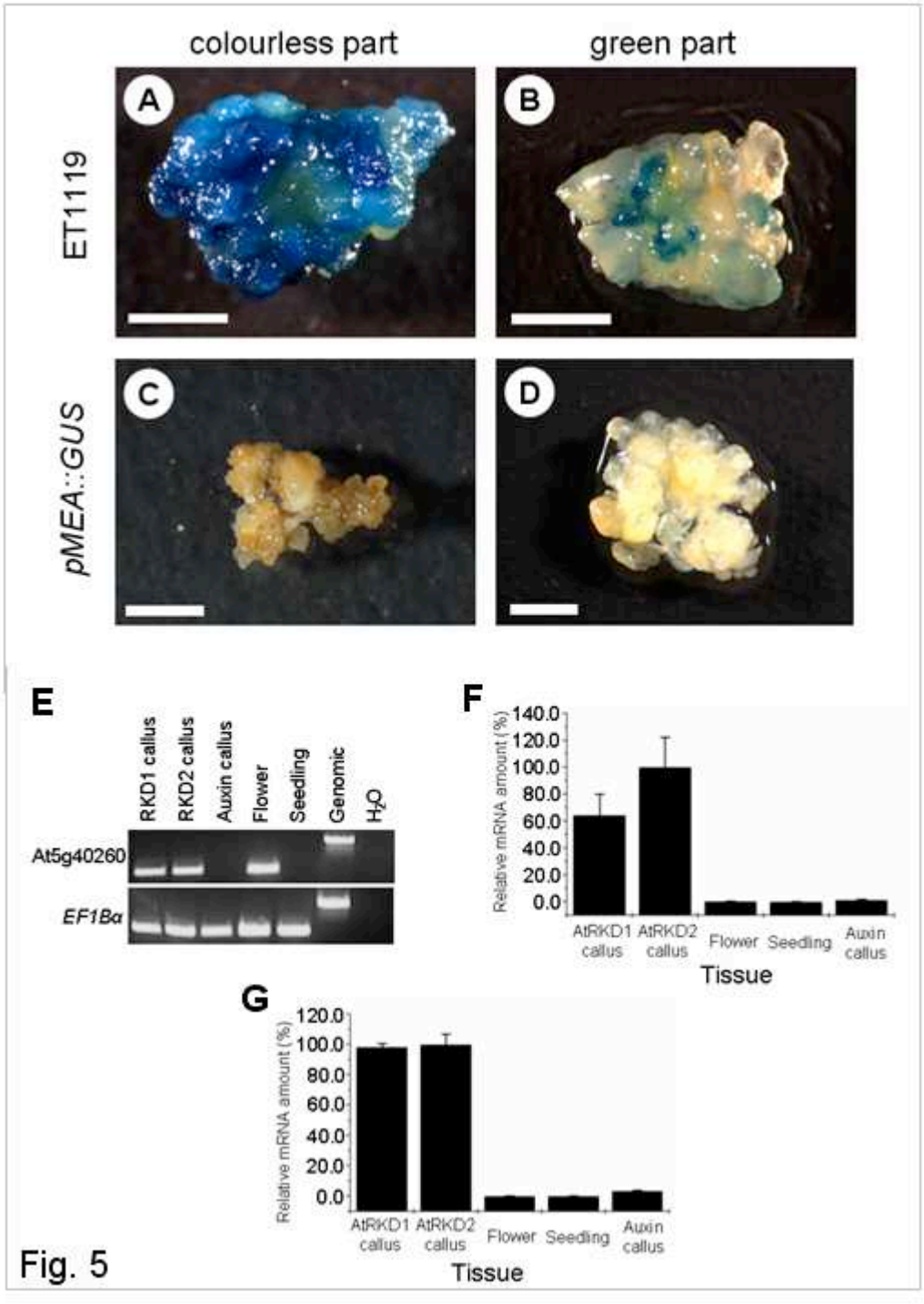

colourless part
green part
ET1119
pMEA::GUS
A
B
C
D
E
RKD1 callus
RKD2 callus
Auxin callus
Flower
Seedling
Genomic
H2O
At5g40260
EF1Bα
F
Relative mRNA amount (%)
140.0
120.0
100.0
80.0
60.0
40.0
20.0
0.0
AtRKD1 callus
AtRKD2 callus
Flower
Seedling
Auxin callus
Tissue
G
Relative mRNA amount (%)
120.0
100.0
80.0
60.0
40.0
20.0
0.0
AtRKD1 callus
AtRKD2 callus
Flower
Seedling
Auxin callus
Tissue

Fig. 5

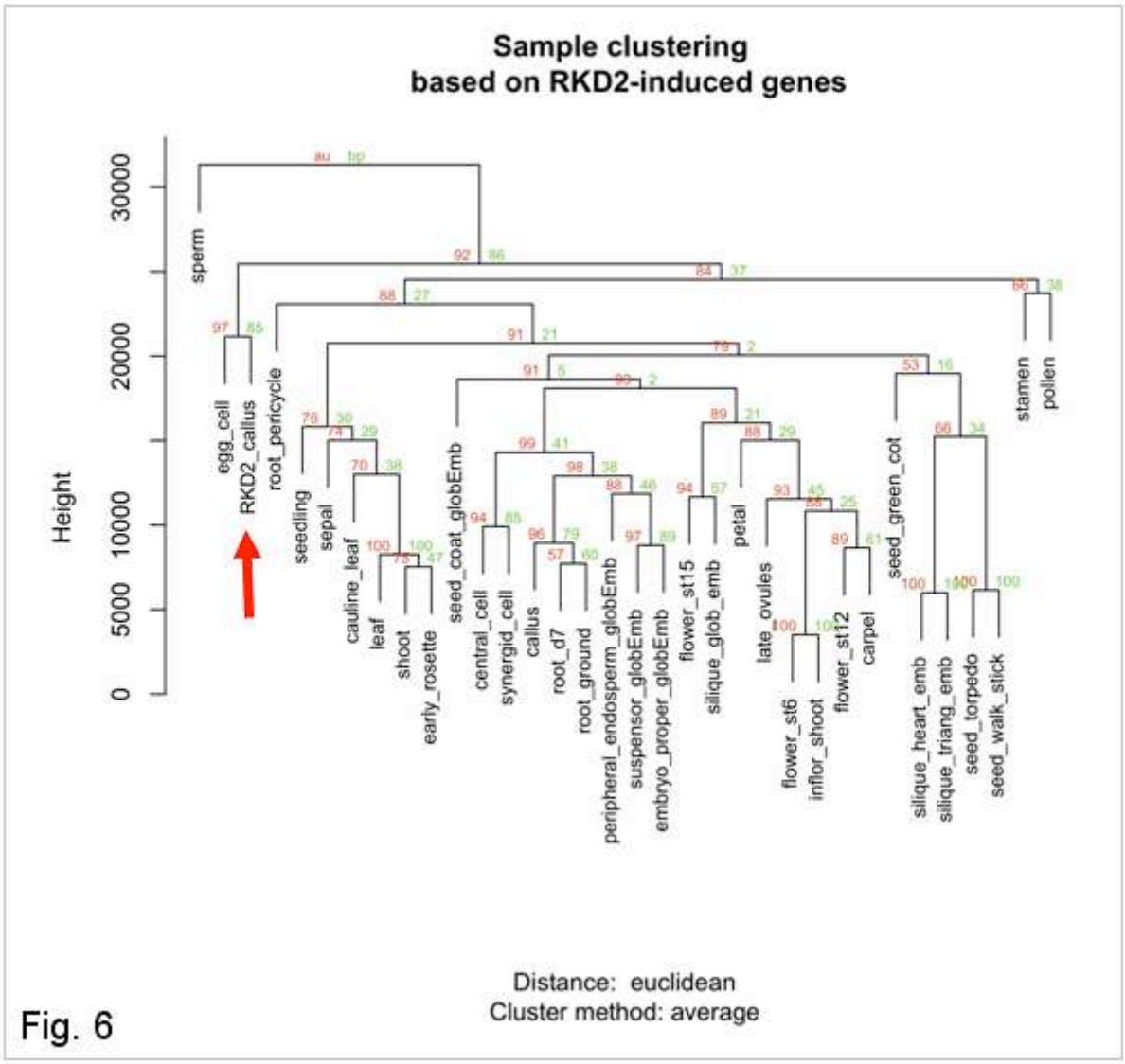
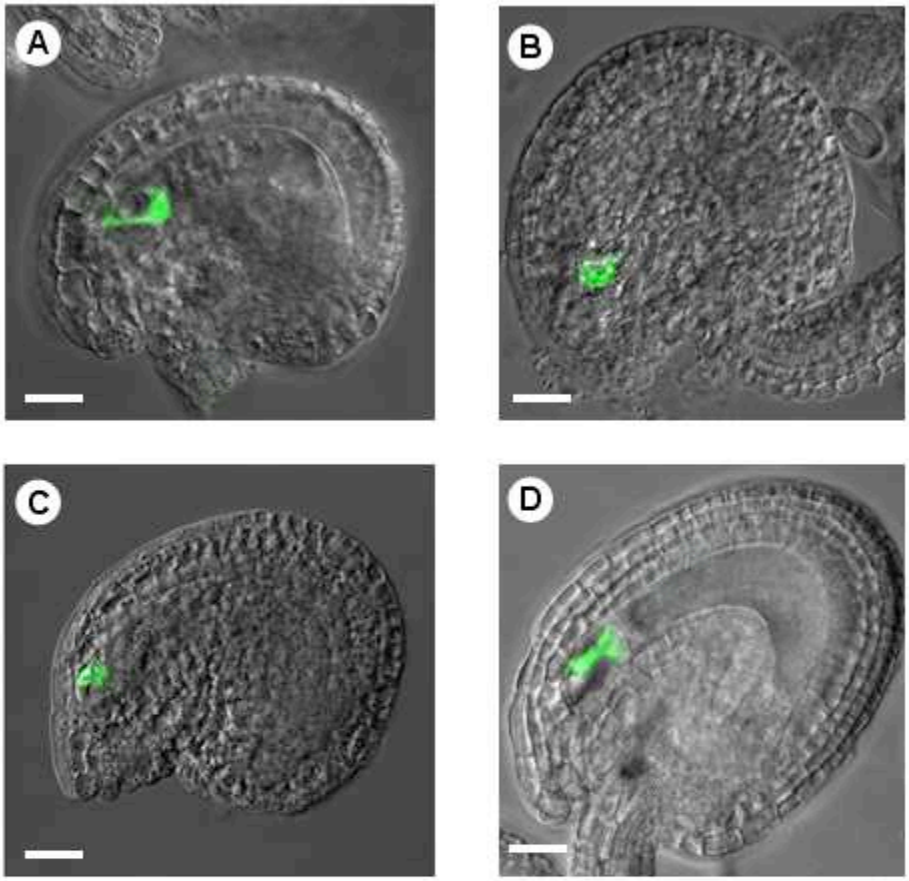

Sample clustering
based on RKD2-induced genes
Height
0
5000
10000
20000
30000
au
bp
sperm
egg_cell
RKD2_callus
root_pericycle
seedling
sepal
cauline_leaf
leaf
shoot
early_rosette
seed_coat_globEmb
central_cell
synergid_cell
callus
root_d7
root_ground
peripheral_endosperm_globEmb
suspensor_globEmb
embryo_proper_globEmb
flower_st15
silique_glob_emb
petal
late_ovules
flower_st6
inflor_shoot
flower_st12
carpel
seed_green_cot
silique_heart_emb
silique_triang_emb
seed_torpedo
seed_walk_stick
stamen
pollen
Distance: euclidean
Cluster method: average

Fig. 6


A
B
C
D

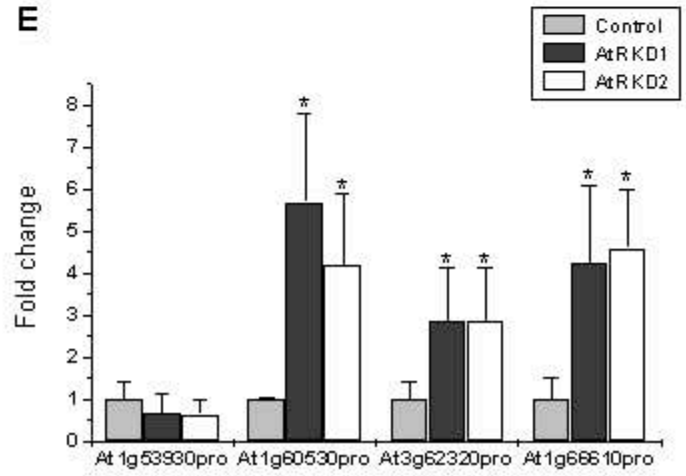

E
Control
AtRKD1
AtRKD2
Fold change
0
1
2
3
4
5
6
7
8
*
At1g53930pro
At1g80530pro
At3g62320pro
At1g66810pro

Fig. 7

Supplementary Figures

Figure S1. Southern blot hybridization of the wheat genome. Genomic DNA was digested overnight by *Eco*RI, *Hin*dIII, *Pst*I, *Xba*I and *Xho*I, separated on a 0.8% agarose gel, blotted, and probed with a 498 bp ^{32}P-labelled *Rsa*I-*Cla*I fragment of *TaRKD*. A size ladder is given for comparison.

Figure S2. Schematic representation of the structures of genomic *TaRKD* genes of *aestivum*-Salmon wheat. Blue boxes represent the coding region, black lines the introns. The grey boxes are the RWP-RK domains. The numbers indicate the lengths of coding regions and introns in bp. The vertical red lines indicate point mutations at the nucleotide level compared to *TaRKD1* and *TaRKD2*. The green scale bar represents 100 bp.

Figure S3. The RWP-RK gene family of *Arabidopsis thaliana*. The RKD subfamily (AtRKD1-5) and the NIN-like subfamily (AtNLP1-9) form two well-separated branches. The tree is based on amino acid sequences and computed by the AlignX software (Vector NTI, Invitrogen, USA).

Figure S4. Schematic representation of *AtRKD1, 2* gene structures and locations of T-DNA insertions. Blue boxes are the coding regions, black lines are the introns and orange lines are upstream regions. The grey boxes are the RKD domains. The red arrows represent the T-DNA insertions. The length of the T-DNA is not to scale. The numbers indicate the lengths of different regions in bp.The green scale bar represents 100 bp.

Figure S5. Subcellular localization of AtRKD-GFP fusion proteins expressed in *Arabidopsis* protoplasts under the control of the double *CaMV35S* promoter. Images were taken by laser-scanning microscopy. White stars indicate the vacuole and white arrowheads label the nucleus. (*A-C*) AtRKD1::GFP; (*D-F*) AtRKD2::GFP; (*G-I*) AtRKD3::GFP; (*J-L*) AtRKD4::GFP; (*M-O*) GFP control. (*A,D,G,J,M*) white light; (*B,E,H,K,N*) UV-light (excitation/emission wavelengths of 488 nm and between 505 nm and 520 nm, respectively); (*C,F,I,L,O*) merged images. Bars, 5 µm

Figure S6. Ectopic expression of the genes (*A*) *AtRKD1* and (*B*) *AtRKD2* under the control of the double *CaMV35S* promoter leads to severe growth distortions, similar to the phenotype of the corresponding AtRKD-GFP constructs. Ploidy level determinations in (*D*) AtRKD1 and (*E*) AtRKD2

over-expressing tissue reveals its 2C level. (*E*) Rosette leaves were used as standard for ploidy analysis (*C*). Bars, 2 mm.

Figure S7. Ectopic expression of *GFP* gene controlled by the double *CaMV35S* promoter (*d35Spro:GFP*). (*A*) white light (*B*) UV light pictures. Scale bars represent 2 mm.

Figure S8. Gamete-specific GUS activity in two marker lines used for super-transformation with AtRKD1-GFP and AtRKD2-GFP under the control of the double *CaMV35S* promoter. Flowers were emasculated and GUS staining was done overnight. (*A*) Egg cell-specific expression in the marker ET1119, and (*B*) central cell-specific expression of *pMEA::GUS* transgene. Bars, 20 μm.

Figure S9. Specific up-regulation of egg-cell enriched genes in RKD2-induced callus. Plot showing proportion of female gametophytically enriched genes amongst genes up-regulated in RKD2-induced callus. Genes were sorted according to fold-change (RKD2-induced callus/control callus), so that lower gene numbers denote genes that are most highly up-regulated in RKD2-induced callus. Among up-regulated genes, the proportion of the gene sets specifically enriched in a given gametophytic cell type when compared to a compendium of tissues and cell types of the plant body is shown for egg cell (red line), synergids (blue line) and central cells (green line). Black lines denote randomly sampled gene lists (of the same size as the marker-list). The graph shows that egg cell markers are significantly enriched amongst over-expressed genes in RKD2-induced callus, as indicated by the p-values of a two-sided Fisher exact test comparing observed and expected gene set proportions at a given fold-change cutoff. Vertical brown lines indicate 3-fold (left), 2-fold (middle) and 1.5-fold (right) up-regulation cutoffs. Total numbers of gametophytically enriched genes are: 222 (egg markers), 138 (central cell markers), and 249 (synergid markers) (Wüst et al. 2010).

Figure S10. Principle component analysis of the log2-signals of RKD2-induced genes demonstrating its close relationship to egg cell.

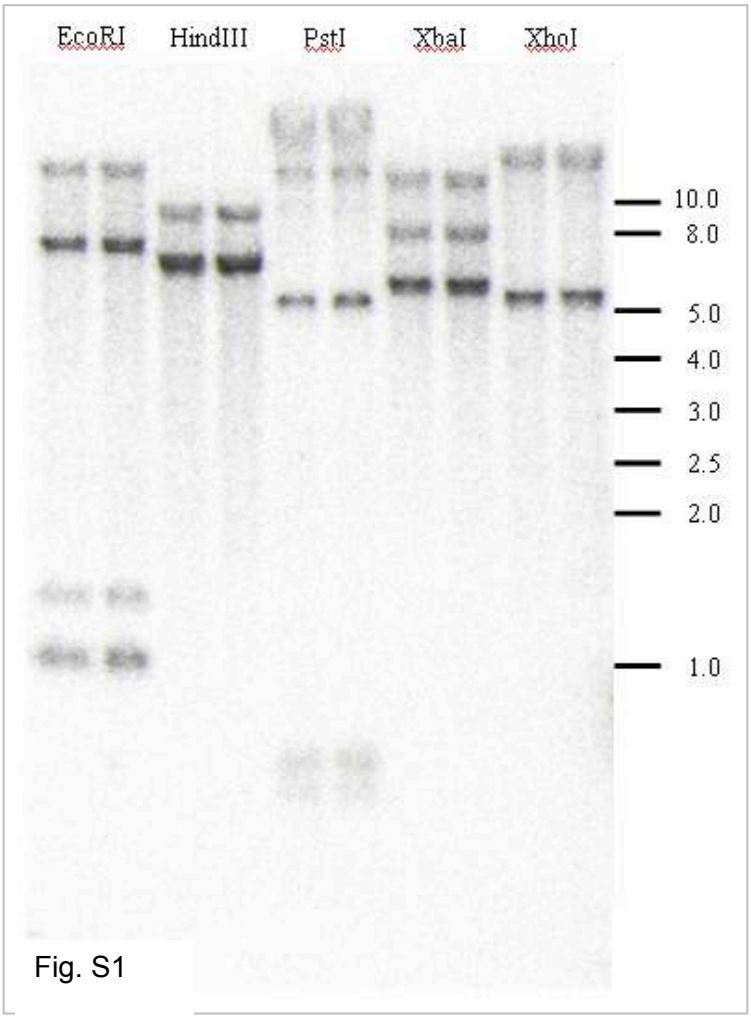


Fig. S1

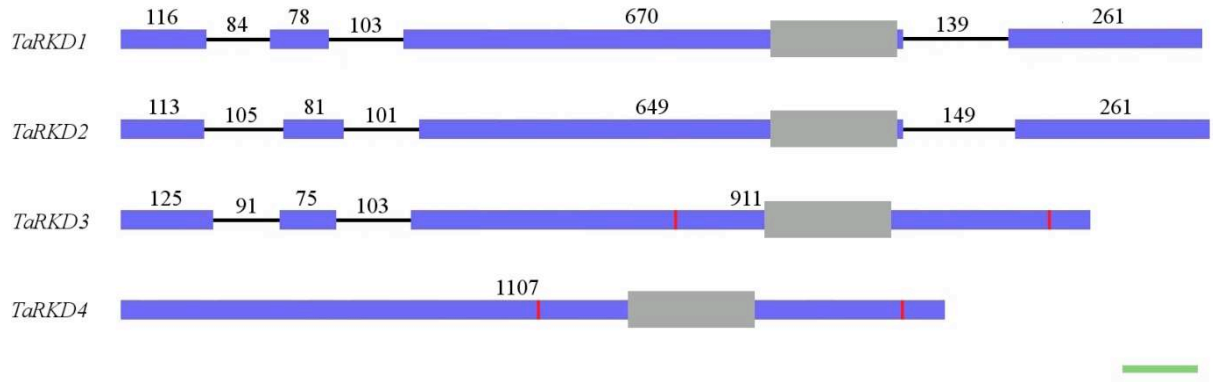

TaRKD1
116
84
78
103
670
139
261
TaRKD2
113
105
81
101
649
149
261
TaRKD3
125
91
75
103
911
TaRKD4
1107

Fig. S2

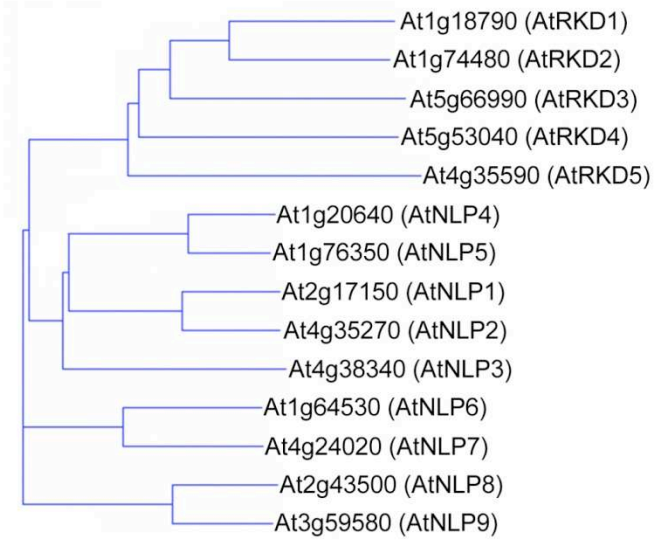

At1g18790 (AtRKD1)
At1g74480 (AtRKD2)
At5g66990 (AtRKD3)
At5g53040 (AtRKD4)
At4g35590 (AtRKD5)
At1g20640 (AtNLP4)
At1g76350 (AtNLP5)
At2g17150 (AtNLP1)
At4g35270 (AtNLP2)
At4g38340 (AtNLP3)
At1g64530 (AtNLP6)
At4g24020 (AtNLP7)
At2g43500 (AtNLP8)
At3g59580 (AtNLP9)

Fig. S3

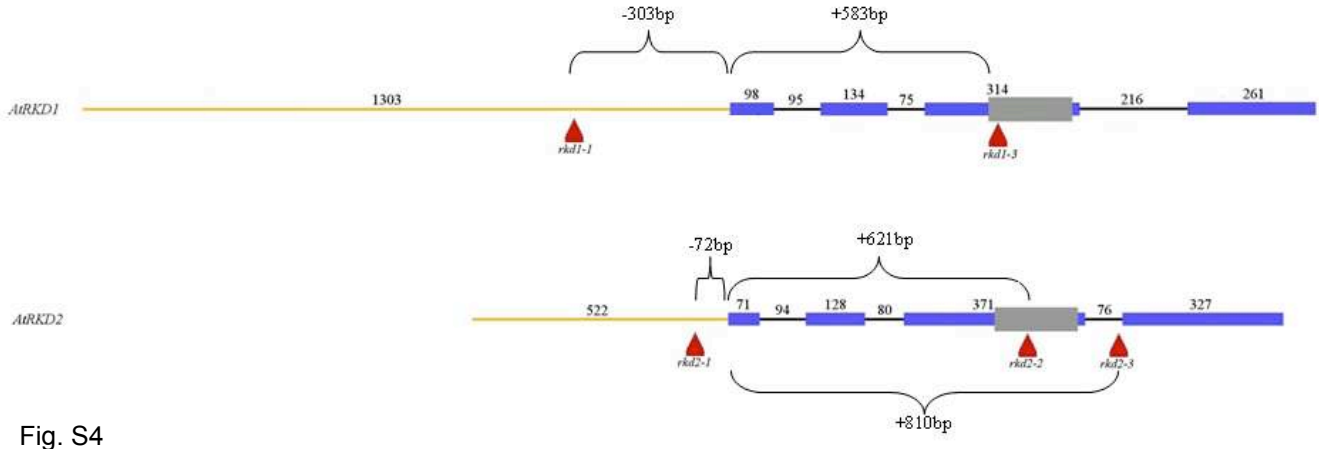

-303bp
+583bp
AtRKD1
1303
98
95
134
75
314
216
261
rkd1-1
rkd1-3
-72bp
+621bp
AtRKD2
522
71
94
128
80
371
76
327
rkd2-1
rkd2-2
rkd2-3
+810bp

Fig. S4

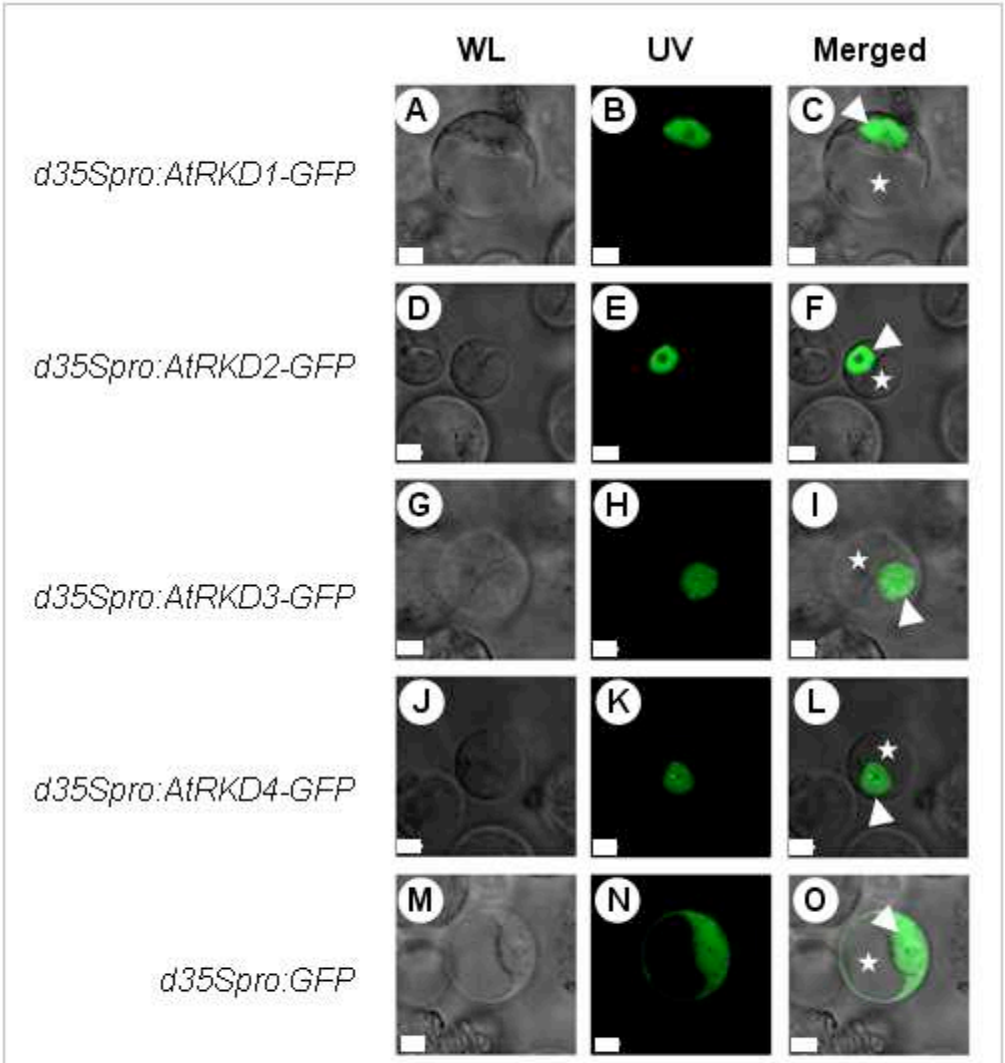

WL
UV
Merged
A
B
C
d35Spro:AtRKD1-GFP
D
E
F
d35Spro:AtRKD2-GFP
G
H
I
d35Spro:AtRKD3-GFP
J
K
L
d35Spro:AtRKD4-GFP
M
N
O
d35Spro:GFP

Fig. S5

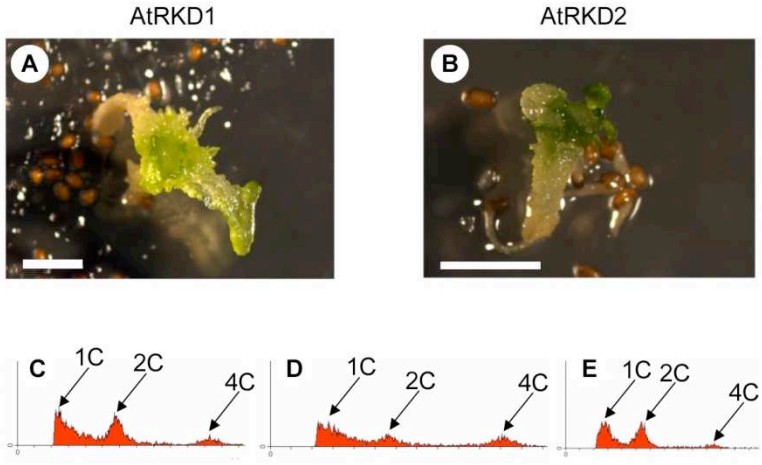


Fig. S6

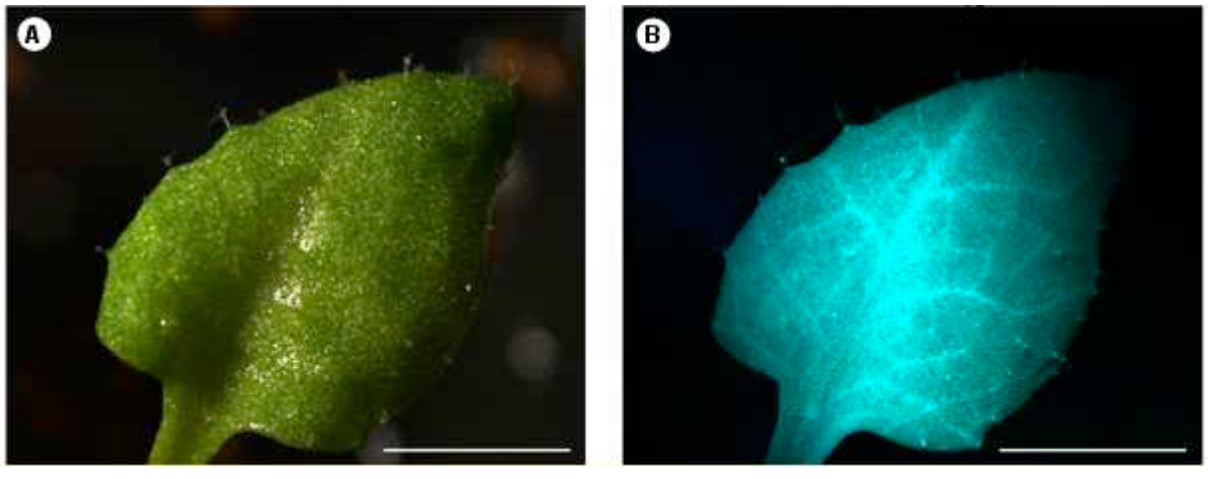


Fig. S7

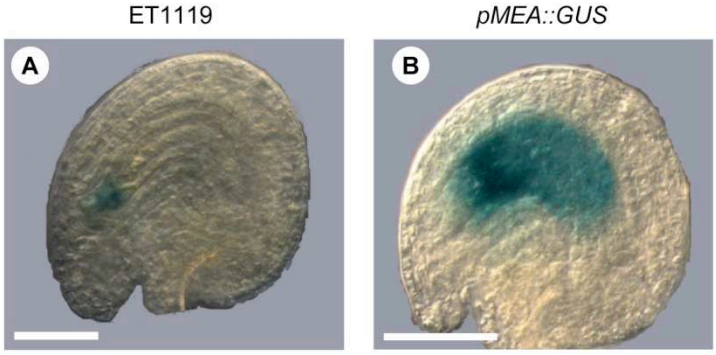

ET1119
pMEA::GUS
A
B

Fig. S8

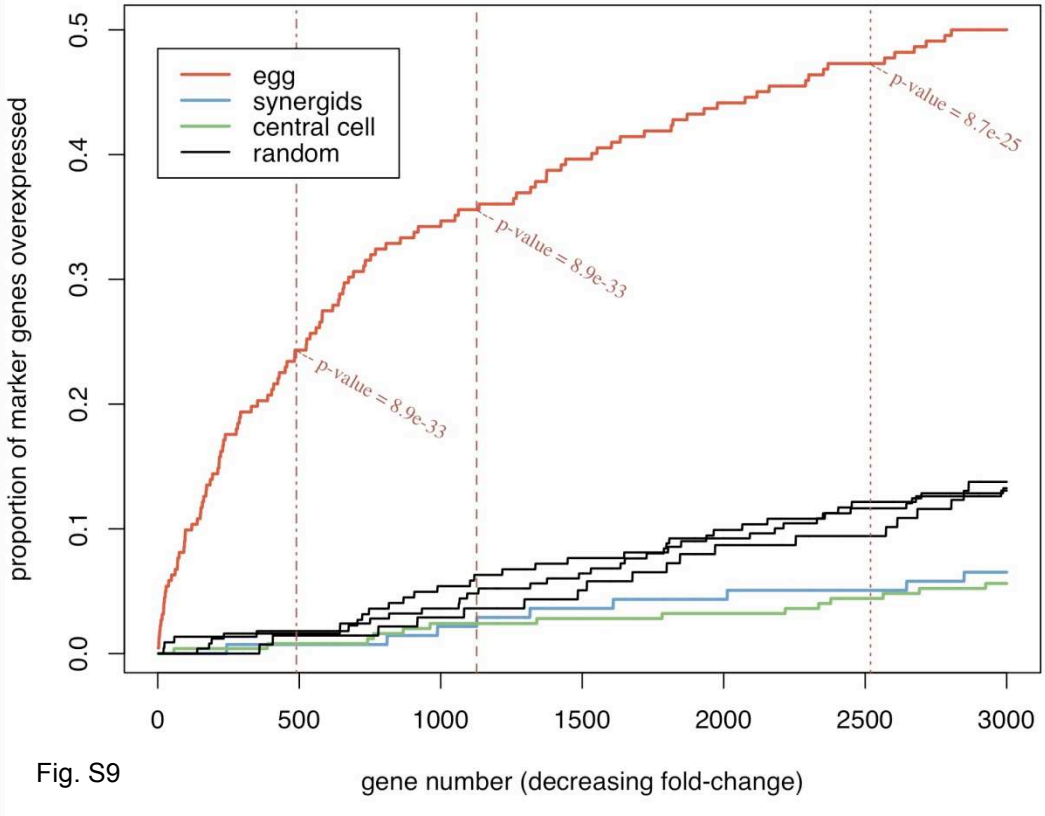

egg
synergids
central cell
random
p-value = 8.9e-33
p-value = 8.9e-33
p-value = 8.7e-25
proportion of marker genes overexpressed
0.0
0.1
0.2
0.3
0.4
0.5
0
500
1000
1500
2000
2500
3000
gene number (decreasing fold-change)

Fig. S9

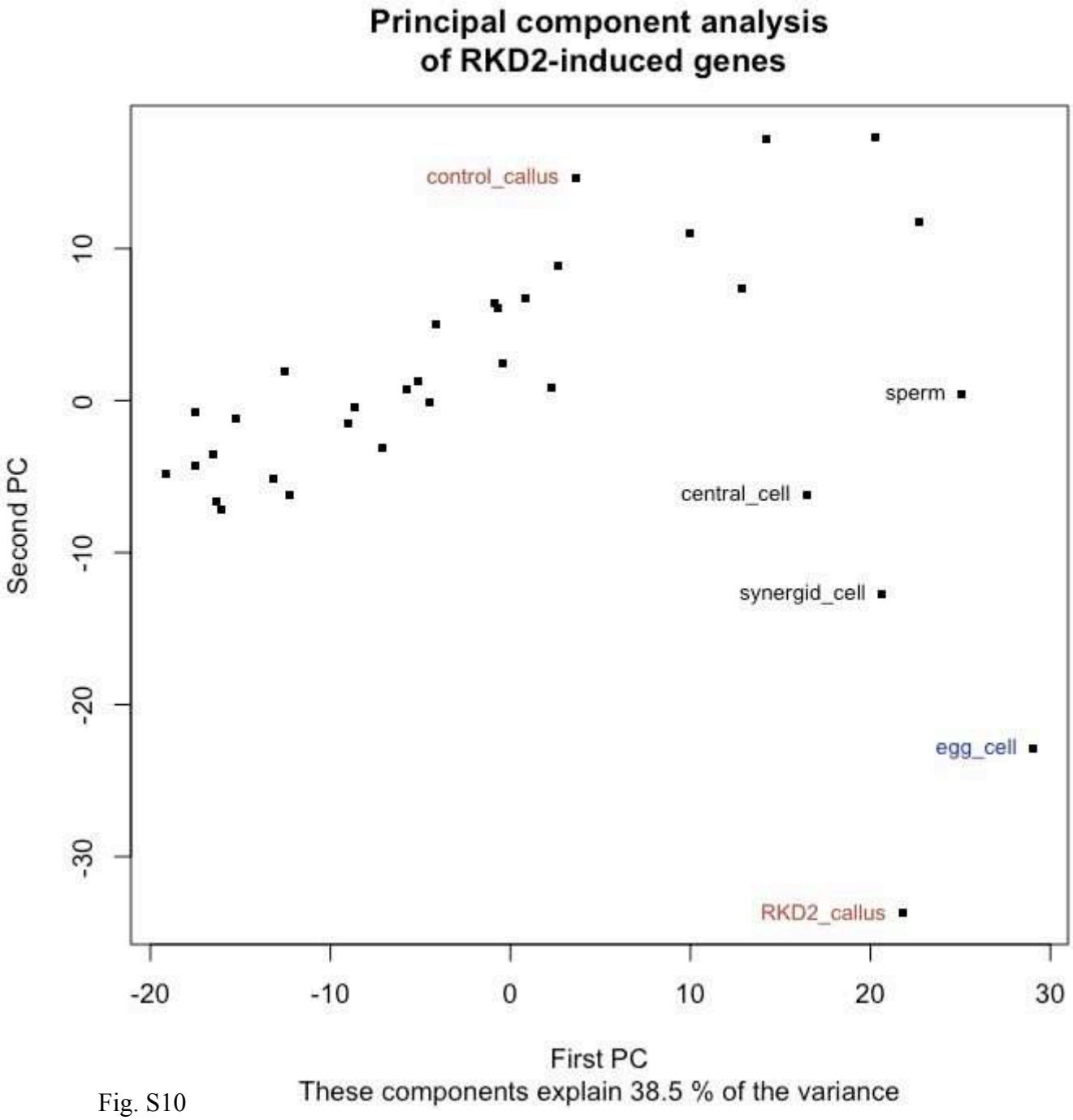

Principal component analysis
of RKD2-induced genes
control_callus
sperm
central_cell
synergid_cell
egg_cell
RKD2_callus
Second PC
First PC
These components explain 38.5 % of the variance
-20
-10
0
10
20
30
-30
-20
-10
0
10

Fig. S10

Supplementary Tables

Table S1. T-DNA mutant alleles in *Arabidopsis thaliana*. Mutant lines were identified from the SALK T-DNA Express database and seeds were received from the Nottingham *Arabidopsis* Stock Centre (NASC) (http://signal.salk.edu (Alonso et al., 2003)) from the GABI-Kat resource (http://www.mpiz-koeln.mpg.de/GABI-Kat/GABI-Kat_homepage.html (Rosso et al., 2003))

Table S2. Identification of AtRKD2 induced genes. Genes were selected based on P value lower than 0.1 and fold change bigger than 3.0. AtRKD2 induced tissue, mean of signal intensities for arrays of AtRKD2 induced tissue. Auxin callus, mean of signal intensities for arrays of auxin callus tissue.

Table S3. Identification of putative egg cell-specific genes from *d35Spro:AtRKD2-GFP* colorless tissue. Genes were selected based on the following criteria P value <0.1 and fold change (FC) 7.0, aginst both controls (auxin induced callus and 14 days old seedlings).

Table S4. Primer names and sequences used in this study.

Table S1.

Gene	Catalogue number/ T-DNA allele	Location of the T-DNA	Genotype	Phenotype
AtRKD1	GABI 522C05 (*rkd1-1*)	5′-UTR	homozygous	none
	SALK 089683 (*rkd1-2*)	coding region	homozygous	none
AtRKD2	SALK 133716 (*rkd2-1*)	5′-UTR	homozygous	none
	GABI 237C07 (*rkd2-2*)	coding region	homozygous	none
	GABI 116G12 (*rkd2-3*)	intron	homozygous	none